

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 12:29:46 : Search time 1930.26 Seconds
(without alignments)
8255.089 Million cell updates/sec

Title: US-09-923-304-1
Perfect score: 722
Sequence: 1 cgcctaagcgcgcgaacatg.....aaggctcttttgacctatt 722

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
- al number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID
Query
Description

1	722	100.0	722	9	D87735
2	695.8	96.5	808	9	BC009294
3	694.4	96.2	721	9	BC019651
4	695	95.0	833	9	BC005134
5	684.8	94.8	760	9	BC000606
6	672.4	93.1	713	9	BC022805
7	668	92.5	753	9	HUMRNAE
8	662.8	91.8	179523	2	AC036104
9	658.8	91.2	191461	2	AC078814
10	647.6	89.1	701	9	HS016738
11	647.6	89.1	175099	2	AC017029
12	595.8	82.5	632	9	HS092982
13	499.4	69.2	212421	2	AC109591
14	485.4	67.2	208632	2	AC073711
15	482.2	66.8	230760	2	AC073714
16	471.2	65.3	715	10	RNRIBOL14
17	441.4	61.1	173493	9	AC093115
18	441.4	61.1	184447	9	AC012519
19	431.6	59.8	190162	2	AC097966
20	413.2	57.2	181238	2	AC094176
21	412.6	57.1	177490	2	AC095342
22	376.4	52.1	573	10	BC008177
23	352	48.8	7729	9	AB061822
24	338	46.8	321708	2	AC087142
25	334.8	46.8	338116	2	AC087159
26	315.4	43.7	176075	9	AC017079
27	312	43.2	185208	30	AC019204
28	312	43.2	187253	2	AC104186
29	302.4	41.9	166585	2	AC108039
30	302.4	41.6	178825	2	AC074251
31	302.4	41.6	178825	2	AC074251
32	302.4	41.6	178825	2	AC074251
33	253.8	36.5	406	5	AB046393
34	252.2	34.9	468	5	AF015691
35	248.4	34.4	457	5	TR031091
36	219.4	30.4	1193	9	BC010079
37	182	25.2	281	11	HS0281P
38	169.2	23.4	638	3	RLA33201
39	151	20.9	167	11	G84294
40	147.8	20.5	283	6	A74662
41	147.8	20.5	283	6	A77641
42	146.4	20.3	184916	2	AC068998
43	146.4	20.3	226199	2	AC084069
44	129.6	18.0	152408	9	HS191318
45	120.2	16.5	163739	2	AC093718

ALIGNMENTS

RESULT	1
D87735	D87735 Homo sapiens mRNA for ribosomal protein L14, complete cds.
LOCUS	D87735 Homo sapiens mRNA for ribosomal protein L14, complete cds.
DEFINITION	D87735 Homo sapiens mRNA for ribosomal protein L14, complete cds.
ACCESSION	D87735
VERSION	D87735.1 GI:1620021
KEYWORDS	ribosomal protein L14.
SOURCE	Homo sapiens neonatal male umbilical cord vein endothelial cell
ORGANISM	Cell_line:HUE4 cdna to mRNA, clone_lib:PCMV-SPORT clone:thue4-4.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	1 (bases 1 to 722)
REFERENCE	Mitsui.Y.
AUTHORS	Submitted (05-SEP-1996) Youji Mitsui, Agency of Industrial Science
TITLE	and Technology, National Institute of Bioscience and
JOURNAL	Human-Technology, Higashi 1-1, Tsukuba Science City, Ibaraki 305.
REFERENCE	Japan (E-mail:ttanaka@is.icc.u-tokai.ac.jp, Tel:+81-298-94-6070,
AUTHORS	Fax:+81-298-94-6095)
TITLE	2 (sites)
REFERENCE	Tanaka.M., Tanaka.T., Harata.M., Suzuki.T. and Mitsui.Y.
AUTHORS	triplet repeat-containing ribosomal protein L14 gene in

	Ctctcgtcgctgctgc	tgc	tgc	tgc	tgc	aagttccaggcaaaaaagatcacccgccga	540
Df	CtGTCGTCTGCGTCTGC	TGCTGCTC	GCTTAAGATTCACGCAAAAAGATCACCCGCCCGGA	540			
Oy	qtaaaaagcgtccagcccagaagtcttcggccagaaaccacagccagaaagccaagcgc						
Dd	GTAATAAAGGGTGCCAGCCCCAGAAGTTCTTGCCAGAAAGCCCACAGCCAGANAAGCAGCGC						
Oy	cgtctccaaaagcgtcagaagggtc	aaaaaaaaaacccccagaaagcacctg	ctcttctaactgtcttgacctg				
Dd	CTGCTCCAAAAAGGTCAGAAAGGGTCAA	AANAGGTCAGGCCCAAGAACACTGCTGCTCCAAGG	550				
Oy	catctggcagaagaataagtagcacaataaataaatgaataaagattcttttgtacctg						
Dd	CATCTGCGCAGANAAGCATNASTGGCANATCATA	NNAAGTANTAANGGTTCTTTTTIGACC TG	720				
Oy	721 tt	722					
Dd	721 TT	722					
RESULT	2						
BC009294	Homo sapiens, clone MGC:16644 IMAGE:4123062, mRNA, complete cds.				PRI 12-JUL-88		
LOCUS	BC009294	808 bp	mRNA linear				
DEFINITION	Homo sapiens, clone MGC:16644 IMAGE:4123062, mRNA, complete cds.						
ACCESSION	BC009294						
VERSION	BC009294.1 GI:14242451						
KEYWORDS	MGC:						
SOURCE	human.						
ORGANISM	Homo sapiens						
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;							
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE	1 (bases 1 to 808)						
AUTHORS	Strausberg,R.						
TITLE	Direct Submission						
JOURNAL	Submitted (12-JUN-2001) National Institutes of Health, Mammali-						
	Gene Collection (MGC), Cancer Genomics Office, National Cancer						
	Institute, 31 Center Drive, Room 1IA03, Bethesda, MD 20892-2595						
	USA						
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov						
COMMENT	Contact: MGC help desk						
	Email: cgapbs@mail.nih.gov						
	Tissue Procurement: AICC						
	cDNA Library Preparation: Rubin Laboratory						
	cDNA Library Arrayed by: The I.M.A.G.E Consortium (LLNL)						
	DNA Sequencing by: National Institutes of Health Intramural						
	Sequencing Center (NISC).						
	Gaithersburg, Maryland.						
	Web site: http://www.nisc.nih.gov/ http://nsc.mgsinhgri.nih.gov						
	Contact: nisc_mgc@nih.gov Shenchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.C., Brinkley,C., Brook, Dietrich,N.L., Guan,X., Gupta,J., Hastrick,L., Karlins,E., Legasp, Lim,M., Maduro,O.L., Masllo,C., Mastrian,S.D., McCloskey,J.C. McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J., Wongson,E.E., Touchman,J.W., Tsurgon,C., Voigt,J.L., Walker,M.R., Zhang,L.-H. and Green,E.D.						
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov/series.html Series: IRAL Plate: 26 Row: C Column: 3 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.						
FEATURES	Location/Qualifiers						
source	1..808 /organism="Homo sapiens"/db_xref="taxon:9606"/clone="MGC:16644 IMAGE:4123062"/tissue_type="Muscle,rhabdomyosarcoma"/clone_lib="NH_MGC_17"/lab_host="DHIOB-R"/note="vector:pOT7"						

QY 11 TCCACATGGTCTCAGGCGCTCTGTAAGATTAGCGGGTGCCTATGCTCCTTGG 70

DB 1 TCCACATGGTCTCAGGCGCTCTGTAAGATTAGCGGGTGCCTATGCTCCTTGG 50

QY 71 ACCTCATGCCGGAATAATTGTCGCGATTGTAGATGTTATTGATCAGAACAGGGCTTGGT 130

DB 61 ACCTCATGCCGGAATAATTGTCGCGATTGTAGATGTTATTGATCAGAACAGGGCTTGGT 120


```

QY 507 caaagctcagaaggtcaaaagctccagcccaagcccaagagcaccctgctcaaggacatctg 566
|||||
Db 605 CAAAAGCTCAGAAAGGTCAGAAAGCTCCAGCCGACAGAAAGCAGCTGCTCAAAAGGTAICTG 564
|||||
QY 567 gcaaaagacataaagtcgaatcataaaaaagtaataaaagttctttaaactatt 722
|||||
Db 565 GCAGAAAGCATAGTGGCAATCATAAAGTAATAAAGTTCTTTTGGCTTIGIT 720
|||||

RESULT 5
LOCUS BC000606 760 bp mRNA linear FRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to ribosomal protein L1b, clone MGC:1544
IMAGE:3345054, mRNA, complete cds.
ACCESSION BC000606
VERSION BC000606.1 GI:12653648
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 760)
Direct Submission
Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DC/DIP
cDNA Library Preparation: Pubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Einkley, C., Brooks, S.,
Dietrich, M.L., Guan, X., Gupta, J., Ho, S.L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Masiello, C., Mastrian, S., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stenrups, S., Thomas, P.J.,
Tiongson, E., Touchman, J.W., Tsurgou, C., Venter, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IMAGE Plate: 5 Row: 9 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched cDNA gi: 4505600.
Location/Qualifiers
1. 760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:1544 IMAGE:3345054"
/tissue_type="Skin, melanotic melanoma"
/clone_lib="NIH_MGC_20"
/lab_host="DH10B-P"
/notes="vector: pOTB7"
45. 695
/codon_start=1
/product="Similar to ribosomal protein L1b"
/protein_id="AAH00506.1"
/db_xref="GI:12653649"
/translation="MVFRPFEVGRVAYVSPGPAQPLVAIVVIDONRALVDGFCIQ
VPQAMFKCMQLDFILKFPFSAHQYVPOAWOKA.LINFEATIPRAKKIEAPEPKA
KMTDFEFKVMKAKMRNRIKNSVKLQKAALEKAEKESGCKTGTAATAAANA
KVPKATIAASKAPACKVPAQKAPAPKALPQVPEKAPAPKASGKKA"

BASE COUNT 241 a 179 c 188 g 152 t
ORIGIN

```

```

Query Match 94.8%; Score 684.8; DB 9; Length 760;
Best Local Similarity 98.1%; Pred. No. 3.6e-137;
Matches 708; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 caccctaacctcccaacatggtgttcacagcgtcttcgaggttgagccgggtgagctatg 60
|||||
Db 28 CGCTTAACGCCGCCAATGGTGTTCAGCGCTTCGTGGAGGTTGGCCGGGTGGCCTATG 8~
|||||
QY 51 tctcctttgacacctatgccggaataattggtccgcatgttagatgttatgtatcagaaca 120
|||||
Db 88 TGCTCTTTGGACCTCAIGCCGGNAATTTGGTCGGATTGTAGATGTTATGTATGATCAACA 117
|||||
QY 121 gggctttggtcgaatgcaccttcactcaagttaggagagcaggccatgcctttcaagtga 130
|||||
Db 148 GGCGTTTGGTCGATGGACCTTGCACTCAAGTAGGAGAGACAGGCCATGCCCTTTCAAGTGA 207
|||||
QY 181 tccagctcaatgatttcacctcaatctctcacagatccaccagagaatattccgac 240
|||||
Db 208 TGCAGCTCACTGATTTTCATCTCAAGTTTCCGCACAGTGCCTCCACCAAGATGTGTCGGAC 257
|||||
QY 241 aagcctggcagagggcagacatcaatacacaataatggcgagccacacgatggcccaaga 300
|||||
Db 258 AAGCCTGGCAGAGGCGAGACATCAATACAAAATGGCGACCCACACGATGGCCCAAGA 327
|||||
QY 301 ttgaagccagagaaagaaagcccaagatcacagattttatcgttttaagtataaagg 340
|||||
Db 328 TTGAGCCAGAGAAAGCAAGCCCAAGATGACAGATTTTGAATCGTTTTTAAAGTTATGAAG 387
|||||
QY 361 caaagaaatgaggaacagaaataatcaagaataaactaaadcttcaaaagcagctc 420
|||||
Db 388 CAAAGAAATGAGGAACAGAAATTAATCAGAAATGAAGTTAAGAAAGTTTCAAAAGCGAGCTC 417
|||||
QY 421 tctcaaaccttctcccaaaagaccttggtactaagggtactgctgctgctgctg 480
|||||
Db 448 TCTGAAAGCTTCTCCAAAAGACCTTGGTACTAAGGGTA-----CTGCTG 425
|||||
QY 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
|||||
Db 496 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 555
|||||
QY 541 gtaaaagctccagcccaagaggttcctgcccagaaagccacagccagcagaaagcagcgc 590
|||||
Db 556 GTAAAAGGCTTCAGCCCAAGAGGTTCTTCCCGAAGAACCCACAGGCCAGAAAGCAGCGC 615
|||||
QY 601 ctgctccaaaagctcagaaggtcaaaaagctccagcccaagaaagcagcctgtcccaagg 650
|||||
Db 616 CTGCTCCAAAAGCTCAGAAAGGTTCAAAAAGCTCCAGCCCAAGAAAGCACCTGTCTCAAAG 675
|||||
QY 661 catctggcaagaaagacataagtggaatacataaaagtaataaagttcttttgaacctg 720
|||||
Db 676 CATCTGGCAAGAAAGCATAAAGTGGCAATCAATAAAGTAATAAAGGTTCTTTTGACCTG 735
|||||

QY 721 tt 722
||
Db 735 TT 737

RESULT 6
LOCUS BC022805 713 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens, clone IMAGE:5426295, mRNA, partial cds.
ACCESSION BC022805
VERSION BC022805.1 GI:18606067
KEYWORDS SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 713)
Direct Submission
AUTHORS Strausberg, R.
TITLE Submitted (04-FEB-2002) National Institutes of Health, Mammalian
JOURNAL

```


OY 200 cctcaagatttctgcacagtgccaccacgaagtatgtctccacagccttgccaggaagcaga 259
 Db 188 CTTCAAGTTTCCGACAGTCCGCCAGCAAGTATGTCCGACAAGCTTGGCAGAAGCCACA 247
 OY 260 catcaatacaaaatggcagccacacagatgggcccagaagaagatgaatcccajagaagaagaa 319
 Db 248 CATCAATAACAAATGGCCAGCCACACAGATGGGCCAAGAAGATTGAAGGCCCAAGAAAGGAA 307
 OY 320 agccaagatgcacagatttttgatcgttttaaaattatgaaggcacaataaaagagaaacagaaacag 379
 Db 308 AGCCAAAGATCAGAGATTITGTGCTTTTAAAGTTATGAAGGGCAAGAAATGAGGAACAG 367
 OY 380 aataatcaagaataaattaaagcttcaaaagcagctctctctctctctctctctctctctctct 439
 Db 368 AATAATCAAGAATAAAGTTAAGAAAGCTTCAAAAGCAGCCTCTCTCTCTCTCTCTCTCTCTCT 427
 OY 440 aaaaagcaccctgactactaagggtactgctctctctctctctctctctctctctctctctctctct 499
 Db 428 AAAAGCACCCTGGTACTAAGGGTAA-----CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
 OY 500 tgcctgctctaaagtccagcaaaaaaagatcacccgcccccagtaaaaagagctccagagccca 559
 Db 492 TCGTCGTCGTAAGTTCAGCAAAAAGATCACCCGCCGCCAGTAAACGTTCCAGGCCCA 541
 OY 560 gaagattctgcccaggaagaccacagggcagagaagcagcagcagcagcagcagcagcagcagc 619
 Db 542 GAAGGTTCTGCCCCAAGAACGACAGCCAGCAAGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
 OY 520 gggtc-aaaaagctccagccagaaagcagcagcagcagcagcagcagcagcagcagcagcagc 579
 Db 502 GGGTCAAAAAGCTCCAGCCAGCAAGCAGCCTGCTCCAAAGGCATCTGGCAGGAAGCAT 561
 OY 579 aagtcgcaatcataaaaaagtaataaaggtttcttttttgacctgttt 22
 Db 562 AAGTGGCAATCAATAAAGTAATAAAGTCTCTTTTTCACCTGTT 205

RESULT 8
 AC026104/C
 LOCUS Homo sapiens chromosome 12 clone RP11-621E11; W. KING DRAFT
 DEFINITION AC026104
 ACCESSION AC026104
 VERSION 11 GI:94383396
 KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Muzny, D.M., Adams, C., Bailey, M., Barberia, J., Blankenburg, K.,
 Bodala, B., Bock, J., Bowie, S., Brooks, A., Bunay, L., Bunac, C.,
 Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
 David, R., Delgado, O., Deshazo, D., Ding, Y., Engh, R., Fashid, N.,
 Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
 Forum-Taney, J., Frantz, P., Ganesh, R., Correll, H., Correll, L.L.,
 Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Jones, M.,
 Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Lea, F., Li, Z.,
 Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Lozano, P., Lu, J.,
 Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, P., Lu, J.,
 Lucier, P., Martin, R., Martinez, C., McLeod, M., M. J. G., Morgan, M.,
 Morris, S., Nash, S., Nelson, A., Nguyen, F., Nguyen, N., Nguyen, S.,
 Oswal, G., Parish, B., Payton, S., Patten, B., Patten, B., Patten, S.,
 Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Su, C.,
 Tabor, P., Taylor, I., Vasquez, L., Vinson, P., V. O., Wabbab, M.,
 Watlington, S., Weinstein, G., Weinstein, I., R., Williamson, A.,
 Worley, K., Wren, J., Wrenford, G., Yu, W., Zhang, X., Nelson, D., and
 Gibbs, R.
 Direct Submission
 JOURNAL
 REFERENCE 2 (bases 1 to 179523)

AUTHORS TITLE JOURNAL

COMMENT

Worley, K.C.
 Direct Submission
 Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 25, 2000 this sequence version replaced gi:8439773.
 ----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HMC
 Center clone name: RP11-621E11
 ----- Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 7% of reads
 Assembly: Dye-terminator Big Dye; 93% of reads
 Consensus quality: 153665 bases at least Q40
 Consensus quality: 167655 bases at least Q30
 Consensus quality: 17310 bases at least Q20
 Estimated insert size: 171802; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation

 NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 26 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 14436: contig of 14436 bp in length
 14437 14536: gap of unknown length
 14537 30933: contig of 16397 bp in length
 30934 31033: gap of unknown length
 31034 47031: contig of 15998 bp in length
 47032 47131: gap of unknown length
 47132 59585: contig of 12454 bp in length
 59586 71143: gap of unknown length
 71144 71243: contig of 11458 bp in length
 71244 82552: gap of unknown length
 82553 91953: contig of 11309 bp in length
 91954 92053: gap of unknown length
 92054 101043: contig of 8990 bp in length
 101044 101143: gap of unknown length
 101144 110670: contig of 9527 bp in length
 110671 110770: gap of unknown length
 110771 119133: contig of 8363 bp in length
 119134 119233: gap of unknown length
 119234 125277: contig of 6044 bp in length
 125278 125377: gap of unknown length
 125378 131392: contig of 6015 bp in length
 131393 131492: gap of unknown length
 131493 138050: contig of 6568 bp in length
 138051 138161: gap of unknown length
 138162 142125: contig of 3965 bp in length
 142126 142225: gap of unknown length
 142226 147719: contig of 5494 bp in length
 147720 147820: gap of unknown length
 147821 151835: contig of 4015 bp in length
 151836 151935: gap of unknown length
 151936 157133: contig of 5198 bp in length
 157134 157233: gap of unknown length
 157234 161670: contig of 4437 bp in length
 161671 161771: gap of unknown length
 161772 164834: contig of 3064 bp in length

Db 96390 CATCIGCAAGAAAGCATAGTGGCAATCATATAAAAGTAATAAAGGTTCTTTTGACCTG 96321
 QY 721 tt 722
 Db 96320 TT 96319

AC078814 191461 bp DNA linear HTG 26-OCT-2001
 Homo sapiens chromosome 12q clone RP11-848D3, *** SEQUENCING IN
 PROGRESS ***, in ordered pieces.
 AC078814
 AC078814.22 GI:13493035
 HTG: HTGS_PHASE2.
 human.

Homo sapiens
 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo:
 1 (bases 1 to 191461)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alstrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Boxie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaey,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
 Cox,C., Cylie,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Recha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
 Garcia,A., Garner,I., Garza,N., Gill,R., Gorrell,J.H., Guetara,W.,
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
 Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
 Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Honsi,F.,
 Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 LcJuseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Marandel,I., Martin,R.,
 Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,
 Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
 Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.I., Morgan,M.,
 Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,M.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenso,S.,
 Oguh,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,R.,
 Peary,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
 Shoshkari,N., Sisson,I., Sodergren,E., Sondaik,I., Sparks,A.,
 Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
 Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,P.,
 Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D.,
 Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
 Washington,C., Watlington,S., Williams,G., Williamson,A.,
 Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
 Zorilla,S., Zucheratpati,R. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 191461)
 Worley,K.C.
 Direct Submission
 Submitted (04-AUG-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 1, 2001 this sequence version replaced gi:13435183.
 INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
 qc-help@bcm.tmc.edu


```
OQ 61 tctcttttgagctcattccggaataattgtctgcgaatttagatgttattgattcagagaaca 120
|||||
Db 123913 TCTCTTTTGGACCTCATGCGGGAATTTGTCGGATTTAGATGTTATTTGATCAGAAACA 123854

OQ 121 gggctttgctgattgagaccttgactcaagtgaaggaacagagccatgcttticaagtgcga 180
|||||
Db 123853 GGGCTTTGGTCGATGGACCTTGACATCAAGTGAAGGACAGAGCCATGCTTTTCAATGGCA 123794

OQ 181 tgcagctcaactgatttcatctcctcaagtcttctgacagctgcccaccagagatgatctccag 240
|||||
Db 123793 TGCAGCTCACTGATTTTCATGCTCAAGTTTCGGACAGTGGCCGCAAAAGTATGTCACAC 123734

OQ 241 aagctgcagagagagagacatcaatacaaaatgggcagagcacacgaaggcccaagaaga 300
|||||
Db 123733 AAGCTTGCAGAGGCGCAGACATCAATAAATAATGGGAGGCGCACAGATGGCCCAAGAGA 123674

OQ 301 tgaagccagagagaagaagccaaagatgacagatttctgacgttttcaagattatgaag 360
|||||
Db 123673 TTGAAGCCAGAGAGAAGGAAGCCCAAGATGACAGATTTTATGCTTTTAAAGTATGAGAG 123614

OQ 361 caaggaataatgaggaacagaataatcaagaatgaagtttaagaagcttcaaaagccagctc 420
|||||
Db 123613 CAAGAAGAACGAGGACAGAGATAATCAAGAAATGAAGTTAAGAGCTTCAAAAGCGACGTC 123554

OQ 421 tctcaaaagcttctccaaaagaccctgctgactgaagtaaggaatactgctctgctgactg 480
|||||
Db 123553 TCTGAAAGCTTCTCCCAAAAAGCCTGTGTACTAAGGGTA-----CTG 123509

OQ 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
|||||
Db 123508 CTGCTGNNNTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 123449

OQ 541 gtaaaaggctcagagccagagagcttctgctgagagagccagagccagagagagagagag 600
|||||
Db 123448 GTAAAAAGGTTTACAGCCAGAGAGTTCTTCCAGAGAGAGCAGAGGCGAGAGAGAGAGAGAG 123389

OQ 601 ctgctcacaagagctcagaagaggtcaaaaagctcagcccaagaagcagctgctcacaag 660
|||||
Db 123388 CTGCTCCAAAAGCTCAGAGAGGTTCAAAAAGCTCCAGCCCAAGAGAGAGCTGCTCCAAAGG 123329

OQ 661 catctgcaagagagagagagagagagagagagagagagagagagagagagagagagagag 720
|||||
Db 123328 CATCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123269

OQ 721 tt 722
||
Db 123268 TT 123267

JUL 10
H:U15739
LOCUS HSU15738 701 bp mRNA linear FRI 11-MAY-1999
DEFINITION Homo sapiens CAG-isl 7 mRNA, complete cds.
ACCESSION U15738 S90520
VERSION U15739.1 GI:508516
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 701)
Aoki,M., Koranyi,L., Riggs,A.C., Wasson,J., Clift,K.C.,
Vaxillaire,M., Froquel,P., Gough,S., Liu,L., Lewis-Keller,H. and
Permutt,M.A.
Identification of trinucleotide repeat-containing genes in human
pancreatic islets
Diabetes 45 (2), 157-164 (1995)
9514166
2 (bases 1 to 701)
Permutt,M.
Direct Submission
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
2
SUBMITTED (31-0CT-1994) M. Alan Permutt, Internal Medicine,
Washington University School of Medicine, 660 S. Euclid, St. Louis.
```

```
MO 63110, USA
COMMENT On May 11, 1998 this sequence version replaced gi:1246101.
FEATURES
Location/Qualifiers
1..701
/organism="Homo sapiens"
/ab_xref="taxon:9606"
/clone="CAG 7"
/tissue_type="pancreatic islets"
7..648
/codon_start=1
/product="CAG-isl 7"
/protein_id="AAC15021.1"
/db_xref="GI:3126984"
/translation="NVRFRFEVGRVAVYFQFHAGKLVAVIDVIDONRALYDGFCTQ
VRPQAMPKCMQMLTDFILKPHSAHQKVRQAWKADINTKWAATRWAKKIEAPERKA
KMTDFDKFKVMAKKMNRRIKNEVKKLQKALLKASPKAPGTGSTIAAAAAAAKVF
AKKIIAASKKATPAOKVPAOKATGOKNAPAPKQKQKAPAKPAPKASGKKA"
repeat=region
452..476
/note="polymorphic region"
/rpt_type=tandem
/rpt_unit=cag
BASE COUNT 225 a 151 c 174 g 140 t
ORIGIN
```

```
Query Match 91.1%; Score 657.4; DB 9; Length 701;
Best Local Similarity 95.9%; Pred. No. 2.8e-131;
Matches 689; Conservative 0; Mismatches 1; Indels 21; Gaps 1;

OQ 12 gcaacatggtgttcaggagcttcgtgaggttgaggcgggtgagcctatgtctcttggga 7;
|||||
Db 1 GCAACATGAGTGTTCAGGGCGCTTCGTGGAGGTTGGCGGGTGGCCATATGTCCTTTTGA 60

OQ 72 cctcatcccggaataatggtcgcgattgtagatgttattgattgacagacagggctttgggc 131
|||||
Db 61 CCTCATGCCGAAATTTGTCGGATTGTAGATGTTATTGATCAGAACAGGCTTTGGTC 120

OQ 132 gatggaccttgactcaagtgaaggaacagagccatgcttcaagtgcagctcact 191
|||||
Db 121 GATGACCTTTCACCTCAAGTGAGGAGACAGAGGATGCTCTTCAAGTGCATGAGCTCAGT 190

OQ 192 gatttcactcgaagtcttcacagatgcccaccagagaatagtccgacaaagcctggcag 251
|||||
Db 181 GATTTCATCTCAAGTTTCGGACAGTGGCCACCAGAGATATGTCGGACAGCCTTGGCAG 240

OQ 252 aaggcagacatcaatacaaaatgggcagccacacgatggggcccaagaagattgaagccaga 311
|||||
Db 241 AAGGCAGACATCAATACAAATGGGCGAGCCACAGATGGCCAGAGATTTGAAGCCAGA 300

OQ 312 gaaagaaagcaagatgacagattttgacttttaagttatgaagcaagaagaaatg 371
|||||
Db 301 GAAGGAAAGCCCAAGATGACAGATTTTGAATGCTTTTAAAGTTATGAAGCAAGAAATG 360

OQ 372 agaacagagaataatcaagaatgaagttaagaagcttcaaaagcagctctcctgaaagct 431
|||||
Db 361 AGAACAGAGAATAATCAAGAAATGAAGTTAAGAGCTTCAAAAGCGAGCTCTCCTGAAGCT 420

OQ 432 tctcccaaaaagcacttgactaaggaaggtactgctgctgctgctgctgctgctgctgct 491
|||||
Db 421 TCTCCCAAAAAGCAGCTGTGTACTAAGGGTA-----CTGCTGCT 450

OQ 492 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 551
|||||
Db 460 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 510

OQ 552 caagccagagaaggttctgctccagagaagccacagagcagagaagagcagcctgctccaaa 511
|||||
Db 520 CCAGCCCAAGAGGTTCTGCTCCAGAGAGCCACAGGCGCAGAGAGAGAGAGAGAGAGAGAG 570

OQ 612 gctcagaagaggttcaaaaagctcagccagagaagcagctgctcgaagcattctgacag 671
|||||
Db 590 GCTCAGAGGGGTCAAAAAGCTCCAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 620
```

```

QY 572 aaagcataaagtgcaatcataaaagataaaggttcctttttgac--gtt 722
      |||||||
Db 540 AAAGCATAAGTGGCAATCATATAAAAGTAATAAAGGTTCTTTTACCTGTT 590
      |||||||

RESULT 11
ACOL17029/c
LOCUS
DEFINITION
Homo sapiens chromosome 12 clone RP11-278E7, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACOL17029
ACOL17029.4 GI:7230953
VERSION
KEYWORDS
HIG: HIGS_PHASE1; HIGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 175099)
Waterston,R.H.
Direct Submission
Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7021919.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/jsc/index.shtml
----- Project Information -----
Center project name: H_NH0278E07
----- Summary Statistics -----
Sequencing vector: M13; 79%
Chemistry: Dye-terminator Big Dye; 21%
Chemistry: Dye-terminator Big Dye; 21% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162555 bases at least Q40
Consensus quality: 166766 bases at least Q30
Consensus quality: 169282 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 173599; sum-of-ctotigs
Quality coverage: 4.18 in Q20 bases; agarose-fp
Quality coverage: 3.90 in Q20 bases; sum-of-ctotigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1365: contig of 1365 bp in length
* 1366
* 1455: gap of unknown length
* 1456
* 3665: contig of 2100 bp in length
* 3666
* 3665: gap of unknown length
* 3666
* 6915: contig of 3251 bp in length
* 6917
* 7016: gap of unknown length
* 7017
* 10547: contig of 3531 bp in length
* 10548
* 10547: gap of unknown length
* 10548
* 19515: contig of 8858 bp in length
* 19516
* 19515: gap of unknown length
* 19516
* 28175: contig of 8560 bp in length
* 28176
* 28175: gap of unknown length
* 28176
* 36507: contig of 8232 bp in length
* 36508
* 36507: gap of unknown length
* 36508
* 45521: contig of 10014 bp in length
* 45522
* 45521: gap of unknown length

```

```

* 45722 57308: contig of 10587 bp in length
* 57209 57408: gap of unknown length
* 57409 67808: contig of 10400 bp in length
* 57809 67908: gap of unknown length
* 67909 81356: contig of 13448 bp in length
* 81357 81356: gap of unknown length
* 81357 94826: contig of 13370 bp in length
* 81457 94826: gap of unknown length
* 94827 111530: contig of 16604 bp in length
* 111531 111631 133064: contig of 21434 bp in length
* 111631 133065 133164: gap of unknown length
* 133065 133165 150963: contig of 17799 bp in length
* 133165 150964 151064: gap of unknown length
* 150964 175099: contig of 24036 bp in length.
* 151064
FEATURES
Location/Qualifiers
Source
1..175099
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-278E7"
1..1365
/misc_feature
/note="assembly_name:Contig1"
1455..3565
/misc_feature
/note="assembly_name:Contig2"
3566..6916
/misc_feature
/note="assembly_name:Contig3"
7017..10547
/misc_feature
/note="assembly_name:Contig4"
clone_end:sp6
vector_side:left
10648..19515
/misc_feature
/note="assembly_name:Contig5"
19516..28175
/misc_feature
/note="assembly_name:Contig6"
28276..36507
/misc_feature
/note="assembly_name:Contig7"
clone_end:t7
vector_side:left
36508..45521
/misc_feature
/note="assembly_name:Contig8"
45522..57308
/misc_feature
/note="assembly_name:Contig9"
57409..67808
/misc_feature
/note="assembly_name:Contig10"
67909..81356
/misc_feature
/note="assembly_name:Contig11"
81457..94826
/misc_feature
/note="assembly_name:Contig12"
94927..111530
/misc_feature
/note="assembly_name:Contig13"
111631..133064
/misc_feature
/note="assembly_name:Contig14"
133165..150963
/misc_feature
/note="assembly_name:Contig15"
151064..175099
/misc_feature
/note="assembly_name:Contig16"
BASE COUNT 52991 a 34531 c 34627 g 51339 t 1511 others
ORIGIN

```

```

Query Match      89.7%   Score 647.6; DB 2;   Length 175099;
Best Local Similarity 95.2%   Pred No. 3.2e-129;
Matches 587; Conservative 0; Mismatches 14; Indels 21; Gaps 1;

```

```

QY 1 cgcctaacgctcccaacatggtgttcaggcgcttcgtgagaggttgccgggtggcctatg 60
      |||||||
Db 128271 CACCTAAGCCCAACCAATGTTGTTTCAGGCGGTTTCGTCAGGTTGCGCGGTTGCTATG 128212
      |||||||
QY 61 tctcttttgacctcatgccgggaaaaatgctgcgcatgtgacatgtttattgatcagaaca 120
      |||||||
Db 128211 TCTCTTTTGACCTCATGCCAGAAAATTGGTCGCGATTGTAGATGTTATTGATCAGAACA 128152
      |||||||

```

```

OY 121 gggcttggtgctgagcttgcactcaagtgagagacagagccatgctttcaagtca 180
Db 128151 GGCTTTGGTTCGATGACCTTGCATCAAGTAGAGAGACAGGCCATGCTTTCAGGTGCA 128092
OY 181 tgaagctactgatttctactctcaagtttctgcacagtgcccccacaggttatgtccac 240
Db 128091 TGCAGTCTACGTGATTTTCATCTCAAGTTTCCGCACAGTCCCGCCAAAGTAIGTCCAAC 128032
OY 241 aagctggcagagaagagcagacatcaatacaaaaatggcagccacacagatggcagagaaga 300
Db 128031 AAGCTTGGCAGAGAGGAGACATCAATACAAAATTTGGCAGCCACACAGA TGGCCAAAGA 127972
OY 301 ttaagccagagaagagaagcagaatgacagagtttgaatgttttaagttatgaag 360
Db 127971 TTGAAGCCACAGAAAGAAAGACAAAGATCACAGATTTTCATCTTTTAAGTAATGAAGG 127912
OY 361 caagagaataggagacagagataatacaagaatgaagttgaagagcttcaaaagggcagctc 420
Db 127911 CAAAGAAAACGAGGACAGACAAATTAATCAAGTAATGAAGTTAAGAAAGCTTAAAGAGGAGCTC 127952
OY 421 tctgaaagcttctcccaaaaagcagcttggtactaaggtactactgtctgctgctgctg 480
Db 127951 TCTGAAGAGCTTCTCCCAAAAAGCAGCTGGTACTACTAAGGGTA----- 127909
OY 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
Db 127909 ---CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127753
OY 541 gtaaaagcctcagccagagaaggtctcctgcacagaaacccagcagagcagagaagcagc 600
Db 127752 GTAAAGAGCTTCAGCCCAAGAGGTTCTCTGCCAGAAAGCCACAGCCAGAAAGGAGGAC 127693
OY 601 ctgctccaaaagctcagaaggttcaaaaagctccagccagagccagagccagagccagagc 660
Db 127692 CTGCTCCAAAAGCTCAGAGAGGTCAAAAAGCTCCAGSCCCAGAAAGCACCTGCTCCAAAGG 127633
OY 661 catctgcaagaagcataagtggaatcatacaaaaagtaataaagttttttacactg 720
Db 127632 CATCTGGCAAGAAAGCATAAAGTGGCAATCATATAAAAGTAATAAAGGTTTITTTGACCTG 127573
OY 721 tt 722
Db 127572 TT 127571

RESULT 12
LOCUS HSU22982
VERSION 1
ORGANISM Homo sapiens
SOURCE human.
REFERENCE
AUTHORS Reddy,P.H., Stockburger,E., Gillevet,P. and Tadde,D.A.
TITLE Mapping and characterization of novel (CAG)n repeat cDNAs from adult human brain derived by the oligo capture method
JOURNAL Genomics 46 (2), 174-182 (1997)
MEDLINE 98096474
REFERENCE
AUTHORS Reddy,P.H., Stockburger,E., Wilderson,J., Ellstrom,J., Gillevet,P. and Tadde,D.A.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1997) Laboratory of Gene Transfer, National Human Genome Research Institute, Room 3A 18, Building 49, Convent Drive, Bethesda, MD 20892, USA
REFERENCE
AUTHORS Reddy,P.H., Stockburger,E., Wilderson,J., Ellstrom,J., Gillevet,P. and Tadde,D.A.
TITLE Direct Submission

```

```

JOURNAL Submitted (12-JAN-1998) Laboratory of Gene Transfer, National Human Genome Research Institute, Room 3A 18, Building 49, Convent Drive, Bethesda, MD 20892, USA
FEATURES
FEATURES
SOURCE
1. 632
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="D1P1B10"
/tissue_type="brain"
/dev_stage="adult"
/complement(381..425)
/rpt_type=tandem
/rpt_unit=CAG
repeat_region
BASE COUNT 200 a 151 c 156 g 124 t 1 others
CPIGIN

Query Match 82.5% Score 595.8 DB 9 Length 632
Best Local Similarity 98.4% Pred. No. 4.9e-118
Matches 522: Conservative 0 Mismatches 8 Indels 2 Gaps 2

OY 95 atttgctgcattgataatgtattgatcaaaagaggttttgctcg-atgagaccttgc 113
Db 1 AATTGCTCGGATTTAGATGTTATTGATCAACAGAGGCTTTGGTCGCTGGACCTTGC 60
OY 144 actcaagtgagagacagagccatgcttttcaagtacatgacactcactgatttcactc 203
Db 61 ACTCAAGTGGAGAGACAGGCCATGCTTTTCAAGTGCATGTCAGCTCACTGATTTCCTC 120
OY 204 agtttctgcagctgccccaccagagatgtctccacagcctgcccagaagacacacac 263
Db 121 AAGTTTCCGCAGTGGCCACCAGAGTGTGTCACCAAGCTGGCAGAGGCACACATC 180
OY 264 atacaagaatggcagccacacacagatggccagaagaattgaagccagagaagaagacc 323
Db 191 AATACAAATGGCAGCCACACAGATGGCCCAAGAGATTGAAGCCAGAGAAAGGAGCC 240
OY 324 aagatgacagattttgctgttttaaagttatgaagcacaagaataatgagaacagaata 383
Db 241 AAGATGACAGATTTTGAGCGTTTANAGTTAAGAGGCAAAAGTAATGGAACAGATA 300
OY 384 atcaagaatgaattgaagaagcttcaaaagcagctctctctgaaagcttct-cccaaaa 442
Db 301 ATCAAGAATGAAGTTAAGAGCTTCAAAAGGCAGCTCTCTCTGAAAGCTTCTCCCAAAA 360
OY 443 accacctgactaagagtaactgctgctgctgctgctgctgctgctgctgctgctgctg 502
Db 361 ACCACCTGGTACTAAGGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
OY 502 tctgctaaagttccagcaaaaagatcacccgcgcagagtaaaagagctccagccagaa 562
Db 421 TCTGCTAAGTTCCAGCAAAAAGATCACCCGCCGAGTTANAGGCTCCAGCCAGAA 480
OY 563 gtttctctccagagaagccacagccacagaagaagcagcctgctccaaaagctcagaagg 622
Db 481 GGTTCCTGCCAGAAAGCCACAGGCCAGAAAGAGCGCGCTGCTCCAAAAGCTCAGAGGG 540
OY 623 tcaaaaagcctccagccagaagaagcactgctccaaaagcactctgcaagaagaagcataagt 582
Db 541 TCCAAAAGCTCCAGCCCAAGAACACCTGCTCCAAAGGCATCTGCCAAGAGCAATAGT 600
OY 683 ggcacataaaaaagtaataaagatttttt 714
Db 601 GGCAATCAATAAAGTAATAAGGGTCTTTTT 632

RESULT 13
LOCUS AC109591/c
DEFINITION Homo sapiens chromosome 4 clone RP11-739K5, WORKING DRAFT SEQUENCE.
ACCESSION AC109591

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

*** nucleic - nucleic search, using sw model

Run on: June 21, 2002, 12:36:51 ; Search time 213.27 Seconds
(without alignments)
5812.408 Million cell updates/sec

Title: US-09-923-304-1

Perfect score: 722

Sequence: 1 cgctaagctgccaacatg.....aaagttcttttgacccgtt 722

Scoring table: IDENTITY.NUC

Gapcp 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

*** total number of hits satisfying chosen parameters: 3:72872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692.4	95.9	888	21 AAF18097	Lung cancer associ
2	567	92.4	709	21 AAC00761	Human secreted pro
3	647.6	89.7	820	21 AAC00765	Human secreted pro
4	636	88.1	730	21 AAC00764	Human secreted pro
5	419.4	58.1	480	21 AAC00763	Human secreted pro
6	413.2	57.2	618	21 AAA16359	Human colon cancer
7	334.8	45.4	409	21 AAS57524	cDNA *200 encoding
8	308.8	42.8	503	22 AAH35325	Human colon cancer
9	295.2	41.0	384	22 AAH98582	Human ESI-derived

c 10	290.2	40.2	539	22	ABA63242	Human foetal liver
c 11	290.2	40.2	539	22	ABA30476	Probe #8942 for ge
c 12	290.2	40.2	539	22	AAK11710	Human brain expres
c 13	290.2	40.2	539	22	AAK37455	Human bone marrow
c 14	290.2	40.2	539	22	AAI18262	Probe #8195 for ge
c 15	290.2	40.2	539	22	AAI43317	Probe #12003 used
c 16	289.6	40.1	360	21	AAC00762	Human secreted pro
c 17	272.8	37.8	504	23	AAS79772	DNA encoding novel
c 18	242.2	33.5	1030	23	AAS83105	DNA encoding novel
c 19	240.8	33.4	401	22	AAK53792	Murine translation
c 20	232.4	32.2	1016	23	AAS79773	DNA encoding novel
c 21	223.8	31.0	395	21	AAC07628	Human secreted pro
c 22	181.8	25.2	220	22	AAC07628	Oryctolagus cunicu
c 23	171	23.7	201	16	AAI15055	Human gene signatu
c 24	145.2	20.2	283	15	AAQ76748	Human genome fragm
c 25	143.5	19.9	335	21	AAC98326	Human colon cancer
c 26	104	14.4	1662	22	ABA15326	Human nervous syst
c 27	99.8	13.8	626	21	AAC50325	Arabidopsis thalia
c 28	99.8	13.8	628	21	AAC35176	Arabidopsis thalia
c 29	98.2	13.6	867	21	AAC51997	Arabidopsis thalia
c 30	97.6	13.5	516	21	AAC47641	Arabidopsis thalia
c 31	97.6	13.5	2232	21	AAC51844	Arabidopsis thalia
c 32	95.2	13.3	602	23	ABLI10777	Drosophila melanog
c 33	95	13.2	619	21	AAC40634	Arabidopsis thalia
c 34	95.6	13.2	674	21	AAC51701	Zea mays DNA fragm
c 35	94.6	13.1	160	22	ABA75690	Human foetal liver
c 36	94.6	13.1	160	22	ABA40284	Probe #18750 for c
c 37	94.6	13.1	160	22	AAK24305	Human brain expres
c 38	94.6	13.1	160	22	AAK50332	Human bone marrow
c 39	94.6	13.1	160	22	AAI27394	Probe #17327 for g
c 40	94.6	13.1	160	22	AAI56302	Probe #24988 used
c 41	91.2	12.6	755	21	AAC51892	Arabidopsis thalia
c 42	80	11.1	485	21	AAC41024	Zea mays DNA fragm
c 43	72.8	10.1	567	21	AAZ29549	HIV codon altered
c 44	54.2	8.9	499	21	AAC51643	Arabidopsis thalia
c 45	53.2	8.8	1037	21	AAA59242	Exons E, C and A c

ALIGNMENTS

RESULT 1
AAFI8097
ID AAFI8097 standard: DNA: 888 BP.
XX AAFI8097:
AC AC
XX AC
XX AC
DI 14-MAR-2001 (first entry)
XX Lung cancer associated polynucleotide sequence SEQ ID 116.
XX Human lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerar;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX Homo sapiens.
XX WO200055180-A2.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000: 2000WO-US05918.
XX FR 12-MAR-1999: 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (EOSE/) ROSEN C A.
XX PI Ruben SM:
XX WPI: 2000-597514/55.
DR

DR P-PSDB; AAB58221.

XX Lung cancer associated gene sequences, referred to as lung cancer
PI antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer.

XX Claim 1: Page 586; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58105 - AAB58549. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective, cytostatic, cardioactive, and
CC immunomodulatory, muscular active general; vulnerability; gastrointestinal
CC general; nephrotropic; antineoplastic; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18432 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterization of the polynucleotide and protein
CC sequences.

XX Sequence 888 BP: 256 A: 216 C: 236 G: 173 T: 7 other:

Query Match 95.9%; Score 692.4; DB 21; Length 888;
Best Local Similarity 99.7%; Pred. No. 6e-165;
Matches 704; Conservative 0; Mismatches 1; Indels 1;

OY 17 catgtgttcaggccttcgtggaggttggcgggtggcgtatgtctcttttgacctca 76
DB 136 catgtgttcaggccttcgtggaggttggcgggtggcgtatgtctcttttgacctca 194
OY 77 tcccgaaattgttcgcgatttagatgttattgacacacagagcttggatcgatgg 136
DB 195 tcccgaaattgttcgcgatttagatgttattgacacacagagcttggatcgatgg 254
OY 137 accttgactcaagtggaggaacagggccatgccttcaagtgcagtcacactgattt 196
DB 255 accttgactcaagtggaggaacagggccatgccttcaagtgcagtcacactgattt 314
OY 197 catctcaagttctgcacagtcgccaccagaagatgtctcgcacacacacacacacac 256
DB 315 catctcaagttctgcacagtcgccaccagaagatgtctcgcacacacacacacacac 374
OY 257 agacatcaatacaaaatgggcagccacacacacacacacacacacacacacacacac 316
DB 375 agacatcaatacaaaatgggcagccacacacacacacacacacacacacacacacac 434
OY 317 gaaagcccaagatgacagattttgatcgttttaagattatcaagccaaagaaatgagaa 376
DB 435 gaaagcccaagatgacagattttgatcgttttaagattatcaagccaaagaaatgagaa 434
OY 377 capataatcaagaataaagtttaagaagcttcaaaagggagcttctctgaaagcttctcc 436
DB 495 cagaataatcaagaataaagtttaagaagcttcaaaagggagcttctctgaaagcttctcc 554
OY 437 caaaaaagcacctgactaaggttactgctgctgctgctgctgctgctgctgctgctgctg 496
DB 555 caaaaaagcacctgactaaggttactgctgctgctgctgctgctgctgctgctgctgctg 514
OY 497 tgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 556
DB 615 tgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 614
OY 557 ccaagaagttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 616
DB 675 ccaagaagttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 734

OY 617 gaaggtcaaaaaagctccagccccaagaagacacctgctccaaaggcatctgcaagaagc 676
DB 735 gaaggtcaaaaaagctccagccccaagaagacacctgctccaaaggcatctgcaagaagc 794
OY 577 ataagtggaatcataaaaaaataataaaaggttctttttgacctgtt 722
DB 795 ataagtggaatcataaaaaaataataaaaggttctttttgacctgtt 840

RESULT 2

ID AAC00761 standard: cDNA: 709 BP.

AC AAC00761:

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 759.

XX Human: 5' EST: expressed sequence tag; secreted protein: cDNA isolation;
KW gene therapy: chromosome mapping: ss.

XX Homo sapiens.

CS EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000: 2000EP-0200610.

XX 26-FEB-1999: 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J. Duclert A. Giordano J;

XX WPI: 2000-500381/45.

XX P-PSDB: AAG00755.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PI obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PI diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX claim 1: SEQ ID 759; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequences. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.

XX Sequence 709 BP: 202 A: 176 C: 183 G: 145 T: 3 other:

Query Match 92.4%; Score 667; DB 21; Length 709;
Best Local Similarity 99.6%; Pred. No. 1.5e-159;
Matches 667; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 cgcctaacgctgcacacatggttctcagcgcttctgaggttgccgggtgacctatg 60
DB 39 cgcctaacgctgcacacatggttctcagcgcttctgaggttgccgggtgacctatg 97
XX 61 tctctttgagacctgcgcgaaaaattggtcgcgatttagatgttattgatcagaaca 120

```

Db 98 tctcctttgagacctatgccggaataattggtcgcgattatgatgtattgatcagaaca 157
QY 121 gggcttttgatcgaatgacaccttgacacaaagtgaagagacagagccatgctttcaagtga 180
Db 158 gggcttttgatcgaatgacaccttgacacaaagtgaagagacagagccatgctttcaagtga 217
QY 181 tgcagctcagctgatttcatctcctcaagtttctcagagtgccaccacagaagtatgtccgac 240
Db 218 tgcagctcagctgatttcatctcctcaagtttctcagagtgccaccacagaagtatgtccgac 277
QY 241 aagcctggcagagggcagacatcaatacaaaaatgggcaagccacacagatgggccaagaaga 300
Db 278 aagcctggcagagggcagacatcaatacaaaaatgggcaagccacacagatgggccaagaaga 337
QY 301 ttgaagccagagaaaggaagccagagatgacagatttgaattttttaaagtattgaagg 360
Db 338 ttgaagccagagaaaggaagccagagatgacagatttgaattttttaaagtattgaagg 397
QY 361 caaagaaatggggaacagaataatacaagaatgaagtttaagaaagcttcaaaaggcagctc 420
Db 398 caaagaaatggggaacagaataatacaagaatgaagtttaagaaagcttcaaaaggcagctc 457
QY 421 tctgaagcttctcccaaaaaagcacctgggtactaaagggtactgctgctgctgctg 480
Db 458 tctgaagcttctcccaaaaaagcacctgggtactaaagggtactgctgctgctgctg 517
QY 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
Db 518 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 577
QY 541 gtaaaaaggctccagccagaggttctgctgctgctgctgctgctgctgctgctgctgctg 600
Db 578 gtaaaaaggctccagccagaggttctgctgctgctgctgctgctgctgctgctgctgctg 637
QY 601 ctgctccaaaagctcagaagggctcaaaagctccaaagctccaaagctccaaagctccaaag 660
Db 638 ctgctcnaaaagctcagaagggctcaaaagctccaaagctccaaagctccaaagctccaaag 697
QY 661 catctgggcaa 670
Db 698 catctgggcaa 707

```

RESULT 3
AF 20765
1c AAC00765 standard: cDNA: 820 BP.

AAC00765:

06-OCT-2000 (first entry)
Human secreted protein 5' EST, SEQ ID NO: 763.

Human: 5' EST: expressed sequence tag: secreted protein: cDNA isolation:
gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PR EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000: 2000EP-0200610.

XX 26-FEB-1993: 99US-0122487.

XX (GSEI) GENSEI.

PI Dumas Milne Edwards J, Duclert A, Giordano J:

DF WBI: 2000-500391/45.

DF P-PSDB: AAG00759.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PI obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PI diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1: SEQ ID 763: 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
mRNAs encoding secreted proteins. An ORF has been identified within the
sequence. The 5' ESTs were prepared from total human RNAs or polyA- RNAs
derived from 30 different tissues. EST sequences usually correspond
mainly to the 3' untranslated region (UTR) of the mRNA because they are
often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
well suited for isolating cDNA sequences derived from the 5' ends of
mRNAs and even in those cases where longer cDNA sequences have been
obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
mRNAs with intact 5' ends and can therefore be used to obtain full length
cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
gene therapy and chromosome mapping procedures. They are used to obtain
upstream regulatory sequences and to design expression and secretion
vectors.

XX Sequence 820 BP: 216 A: 212 C: 222 G: 163 T: 7 Other:
Query: Match 89.7% Score 647.6: DB 21: Length 820:
Best Local Similarity 98.9%: Pred. No. 1.4e-154:
Matches 647: Conservative 3: Mismatches 4: Indels 0: Gaps 0:

```

QY 17 catgtgttcaaggcgtctctgaggggtggccgggtggccttatgtctctcttgaacctca 76
Db 165 catgtgttcaaggcgtctctgaggggtggccgggtggccttatgtctctcttgaacctca 224
QY 77 tgcgggaaaaattgctgcgaattgttagattgttatgcagaacaggcgtttggctcagtg 176
Db 225 tgcgggaaaaattgctgcgaattgttagattgttatgcagaacaggcgtttggctcagtg 284
QY 137 acctgtcactcaagtgaagacagagccatcctttcaagtgcagtcagctcactgattt 195
Db 285 acctgtcactcaagtgaagacagagccatcctttcaagtgcagtcagctcactgattt 344
QY 167 catctcaagtcttctgcacagtgccaccagagaatgtctccgacaagcctggcagaagcc 256
Db 345 catctcaagtcttctgcacagtgccaccagagaatgtctccgacaagcctggcagaagcc 404
QY 257 agacatcaatacaaaaatggcagccacacagatgggccaagaagattgaagccagagaag 316
Db 405 agacatcaatacaaaaatggcagccacacagatgggccaagaagattgaagccagagaag 464
QY 317 gaaagccaagatgacagattttgatcgttttaaaagtattgaaggcaaaagattgaagccaagaag 464
Db 465 gaaagccaagatgmnkgattttgatcgttttaaaagtattgaaggcaaaagattgaagccaagaag 524
QY 377 cagaataatacaaaaatgaagttaagaagcttcaaaaggcagctctctgaaagcttctcc 436
Db 525 cagaataatacaaaaatgaagttaagaagcttcaaaaggcagctctctgaaagcttctcc 584
QY 437 caaaaaagcaccttggtactaaggggtactgctgctgctgctgctgctgctgctgctgctg 496
Db 585 caaaaaagcaccttggtactaaggggtactgctgctgctgctgctgctgctgctgctgctg 644
QY 497 tgcgtgctgcttaaaagttccagcaaaaaagatcacccgcgcgcgcgcgcgcgcgcgcgcgc 555
Db 645 tgcgtgctgcttaaaagttccagcaaaaaagatcacccgcgcgcgcgcgcgcgcgcgcgcgc 704
QY 557 ccagaaggttctgctccagaaaagccacagagcagagaagcagcgtctgctcccaaaagctca 616
Db 705 ccagaaggttctgctccagaaaagccacagagcagagaagcagcgtctgctcccaaaagctca 764
QY 617 gaagggtcaaaaagctccagcccagaaaagcacctgctcccaaaagcattctggcaa 670
Db 765 gaagggtcaaaaagctccagcccagaaaagcacctgctcccaaaagcattctggcaa 818

```


AA557524	
ID	AA557524 standard; cDNA: 409 BP.
XX	
AC	AA557524;
XX	
DI	13-FEB-2002 (first entry)
XX	
XX	cDNA #200 encoding portion of a human colon tumour protein.
DE	
XX	
FW	Human: colon tumour protein; colon cancer; gene therapy: cytostatic; ss.
XX	
OS	Homo sapiens.
XX	
PN	WQ200173027-A2.
XX	
PD	04-OCT-2001.
XX	
PF	22-MAR-2001; 2001WO-US09246.
XX	
PF	24-MAR-2000; 2000US-191597P.
XX	
PF	04-MAY-2000; 2000US-202024P.
XX	
PF	05-MAY-2000; 2000US-202189P.
XX	
PF	(CORI-) CORIXA CORP.
XX	
PF	Meagher MJ, Xu J, King GE:
XX	
PF	WPI: 2001-611627/70.
XX	
PI	New colon tumour proteins and related nucleic acid, useful for
XX	treatment, prevention, diagnosis and monitoring of cancer -
PI	
XX	Claim 4; Page 91: 299pp; English.
PS	
XX	
XX	The present invention relates to the isolation of novel cDNA sequences
XX	encoding for at least an immunogenic portion of human colon tumour
CC	proteins. The sequences of the invention are useful in pharmaceutical
CC	compositions and vaccines for the prevention and treatment of cancers
CC	such as colon cancer. They are also useful for the diagnosis and
CC	monitoring of such cancers. Antibodies to the colon tumour proteins
CC	and antigen presenting cells that express polynucleotides encoding
CC	colon tumour proteins can be used to inhibit the development of
CC	cancers. T-cells that react specifically with colon tumour proteins
CC	are useful for removing tumour cells from samples (e.g. blood) and
CC	for cancer treatment. The polynucleotides sequences are also useful in
CC	gene therapy. AA557325-AA55880 represent the cDNA sequences of the
CC	invention that encode for portions of human colon tumour proteins.
XX	
XX	Sequence 409 BP: 100 A: 100 C: 116 G: 91 T: 2 other:

Query Match	46.4%	Score 334.8:	DB 23:	Length 403:
Best Local Similarity	98.8%	Pred. No. 2.4e-75:		
Matches 336:	Conservative	0:	Mismatches	4:
			Indels	0:
			Gaps	0:
Oy	8	Cgctgcacacatgatttcaggcgcttcctgtagagattgcccggatggcctatgtctcctt	67	
Dp	70	cgcgcacacatgggttcaggcgcttcctgtagagattgcccggatggcctatgtctcctt	129	
Oy	68	tggacctcatccgcggaaaattggtcgcatgtgacatgtattatcagacacaggcttt	127	
Dp	130	tggacctcatccgcggaaaattggtcgcatgtgacatgtattatcagacacaggcttt	189	
Oy	128	ggtcgatggaccttgcactcaagtggagagacagccatgcctttcaagtgcacgact	197	
Dp	190	ggtcgatggaccttgcactcaagtggagagacagccatgcctttcaagtgcacgact	249	
Oy	188	cactgatttcacatccacgatttctgcacagtggccacacaaagtatctcgacacagcctg	247	
Dp	250	cactgatttcacatccacgatttctgcacagtggccacacaaagtatctcgacacagcctg	309	
Oy	248	gcagaaggcagacatcattacaaaattggcagccacacacgattggcccagagattgaagc	307	

D	B	310	gcagaaagccagacatcnaatacaaaatgggcagccacacgatggccagaaagatigaagc	369
Q	Y	308	cagaaaggaagaaagcccaagatgacagatatttgcgtttt	347
D	B	370	cagaaaggaagaaagcccaagatgacagatatttgcgtttt	409
RESULT 8				
AAH35325				
ID	AAH35325=standard; cDNA; 503 BP.			
XX	AC			
XX	AC	AAH35325;		
XX	DT	03-SEP-2001 (first entry)		
XX	XX			
DE	XX	Human colon cancer antigen encoding cDNA SEQ ID NO:2407.		
XX	XX			
KW	XX	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	KW	colorectal carcinoma; ss.		
XX	XX			
CS	XX	Homo sapiens.		
XX	XX			
FN	XX	WC200122920-A2.		
PD	XX			
XX	XX	05-APR-2001.		
XX	XX			
PF	XX	28-SEP-2000; 2000NO-US26524.		
XX	XX			
PR	XX	29-SEP-1999; 99US-0157137.		
PR	XX	03-NOV-1999; 99US-0163280.		
XX	XX			
PA	XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	XX			
PI	XX	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX	XX			
DR	XX	WPI: 2001-235357/24.		
DR	XX	P-PSDB: AAG75920.		
XX	XX			
XX	XX	Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -		
PS	XX	Claim 1: Page 3958; 9803pp; English.		
XX	XX			
CC	CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing P. Additionally, N may be used to produce the colon cancer-associated ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention.		
CC	CC	N.B. Pages 665 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.		
CC	CC			
XX	XX	Sequence 503 BP; 182 A; 120 C; 96 G; 96 T; 9 other;		

```

Query: Match 42.8% Score 308.8; DB 22: Length 503;
Best Local Similarity 94.1% Pred. No. le-68;
Matches 334; Conservative 0; Mismatches 11; Indels 11; Gaps 3;

Q: 254 ggcagacatcaatacacaatgggcagccacacgcatgggccaaagaattgaagccagaga 313
|||||
Db 13 ggcagacatcaatacacaatgggcagccacacgcatgggccaaagaattgaagccagaga 72
|||||

```



```

OJ 188 cactgatttcctcaagtcttcgcacagtgcacca--ccagaagtatgccacaagacc 245
Db 237 CACTGACTTATTCCTCAAGTTCACCACCGAGTCCGCCACCCAGAGGCATGCCAAGAAGGCC 178
OJ 246 tggcaaaaggcagacatcaatacaaaatggggcagccacacacatggggccaaagaagattgaa 305
Db 177 TGCAGAGAGCGACACCATCATACAAATGGACAGGACACAAATGGGCGCAAGAAGAGTGAA 118
OJ 306 gccagagaaaggaaagcccaagatgacagattttgatcgttttaagtattgaagcacaag 365
Db 117 GCCAAGAAGAGGGCAAGCCAAAATCATAGATTTTGTATCAATTT--AGTCATGAGGCGAAG 50
OJ 356 aaaaatgagaacaggaataatacaagaatgaagttaaaagacttcaaaagcagctctct 424
Lb 59 AAAAATGAGGACAGAGATGATCAAGAAATTAATTAAGAAGCTTCAAAAGCTGCTGCTC 1

RESULTS
AA118262/C
ID AA118262 standard; DNA: 539 BP.

AA118262:
12-OCT-2001 (first entry)
Probe #195 for gene expression analysis in human cervical cell sample.
Probe: human: microarray: gene expression; cervical epithelial cell;
cervical cancer; ss.
Homo sapiens.
WO200157278-A2.
O9-AUG-2001.
30-JAN-2001: 2001WO-US00670.
04-FEB-2000: 2000US-0189312.
26-MAY-2000: 2000US-0207456.
30-JUN-2000: 2000US-0408408.
03-AUG-2000: 2000US-0532395.
21-SEP-2000: 2000US-0234487.
27-SEP-2000: 2000US-0236359.
04-OCT-2000: 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488201/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells.
Claim 25; SEQ ID No 8195; 487pp; English.
The present invention relates to human single exon nucleic acid probes
(SENPs). The present sequence is one such probe. The SENPs are derived
from human HeLa cells. The SENPs can be used to produce a single exon
microarray, which can be used for measuring human gene expression in a
sample derived from human cervical epithelial cells. By measuring gene
expression, the probes are therefore useful in grading and/or staging
of diseases of the cervix, notably cervical cancer.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 539 BP: 130 A: 142 C: 116 G: 151 T: 0 other:

```

Query Match	40.2%	Score 290.2	DB 22	Length 539
Best Local Similarity	85.7%	Pred. No. 54e-64		

	Matches	359;	Conservative	0;	Mismatches	53;	Indels	7;	Gaps	3;
Oy	e	cgcctgccacaatgggttcacgagcgttcctcgtagagggttgccggtgacctatgtctcctt	57							
Dd	414	CGCCACCAACAATGGTGTTCAGGCACATTTGTGGAGGTTGGCCAAGTGCTTACATCTC---	356							
Oy	58	tgaacctcatcccggaaaaattggtcgcgattgtagattgttatgatcagaacaggcgttt	127							
Dd	357	TGGACCTCATCTGGAAAGCTGGTTGGCGATTATAGATTGTTTCATCAGAACAGGSCITTT	298							
Oy	128	ggtcgatggacttcacacaaagtgaagagagcaagccaatgccctttaagtgcatacagct	187							
Dd	297	GGIITGGAGACTTGCCTCAGCTTAGAGAGACAGGCCATGCCATTTAAATGGCAATGACGT	238							
Oy	188	cactgatctcatcctcaagtcttcgcacagtgcacca--ccagaaglatgtcccacaagcc	245							
Dd	237	CACCTGACTTTAICTCAAGTCCCACCAGTGCCTCCACCCAGAGGCGATGCCCAAGAGCC	178							
Oy	246	tggcagaagccagacatcaatacaaaaatggcgagccacacgcatggccccaaagattgaa	305							
Dd	177	TGGCAGAAGSCAACCATCATATACAAAATGGACAGGCGCACACATGGGCCAAGAGAGTGAA	118							
Oy	306	gccagagaaaggaagccaagatgcagagatttgttcgttttaaaqgtatgaaggcacaag	365							
Dd	117	GCCAAAGAGGGGAAGCCAAAATGATAGATTTTGATCAITTT--AGTCATGAAGGCCAAG	50							
Oy	366	aaatgaggaacagaataatcaagaatatgaagtttaagaagcttcaaaagcgactctcct	424							
Dd	59	AAAAATGASGACAGATGATCAAGAATTAAAITTAAGAAGCTTCATAAAGGCTGCTGCTCI	1							
RESULT 15										
AA143317/C										
ID	AA143317 standard; DNA: 539 BP.									
XX	AA143317;									
AC										
XX										
DI	17-OCT-2001 (first entry)									
DE										
XX										
XX	Probe #12003 used to measure gene expression in human placenta sample.									
KK	Probe: microarray; human; placenta; antenatal diagnosis;									
KK	genetic disorder; ss.									
XX										
OS	Homo sapiens.									
XX										
PH	W0200157272-A2.									
PD	09-AUG-2001.									
PF										
XX	30-JAN-2001; 2001WO-US00663.									
PR										
PR	04-FEB-2000; 2000US-0180312.									
PFR	26-MAY-2000; 2000US-0207456.									
PFR	30-JUN-2000; 2000US-0608408.									
PFR	03-AUG-2000; 2000US-0632365.									
PFR	21-SEP-2000; 2000US-0234687.									
PFR	27-SEP-2000; 2000US-0236359.									
PFP	04-OCT-2000; 2000GB-0024253.									
XX										
RPA	(MOLE-) MOLECULAR DYNAMICS INC.									
XX										
FPI	Penn SQ. Hanzel DK. Chen W. Rank DR.									
XX										
DR	KFI: 2001-488897/53.									
XX										
PFI	Human genome-derived single exon nucleic acid probes useful for									
XX	analysing gene expression in human placenta -									
SS	claim 25; SEQ ID No 12003; 654pp; English.									
XX										
XX	The present invention relates to single exon nucleic acid probes (SENPs).									
XX	The present sequence is one such probe. The probes are usefulfor									

1

CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

[illegible]

Search completed: June 21, 2002, 14:24:21
Job time: 5450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2002, 11:55:00 : Search time 1700.22 seconds
(without alignments)
5731.157 Million cell updates/sec

Title: US-09-923-304-1

Perfect score: 722

Sequence: 1 cgctaagctgcaacatg.....aaggtctttttgaactgtt 722

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	698.2	96.7	755	10	BG756624
2	693.2	96.0	822	10	BG763626
3	691.8	95.8	730	10	BG109146
4	690.2	95.6	901	10	BG748004
5	688.4	95.3	797	10	BM083723
6	688.0	94.3	728	9	AU123120
7	680.4	94.2	818	9	AL531631
8	679.8	94.2	723	10	BM472466
9	679	94.0	730	10	BG396368
10	677.4	93.8	746	10	B1914072
11	675.2	93.5	732	10	BG260421
12	674.4	93.4	705	10	BM043710
13	673.4	93.3	712	9	AL528184
14	671.4	93.0	1116	10	BM480399
15	670	92.8	724	10	BG437749
16	668.4	92.6	711	10	BM478504
17	668.2	92.5	732	10	BM009159

18	555.6	92.3	673	10	BG774138
19	556.2	92.3	752	10	BM423398
20	554	92.0	723	10	BM019074
21	564	92.0	845	10	BG473548
22	563	91.8	769	10	B1856897
23	561.8	91.7	716	10	BG764388
24	561.2	91.6	728	10	BG485150
25	560.8	91.5	749	9	AL583702
26	560.6	91.5	807	9	AL518218
27	560.4	91.5	708	9	AL526194
28	559.8	91.4	778	10	B1252458
29	559	91.3	748	10	BG282209
30	558.8	91.2	715	10	BM471385
31	558	91.1	876	10	BG489161
32	556.2	90.9	808	10	BG751023
33	555.4	90.8	1166	10	BM466808
34	554.4	90.6	598	10	B1092706
35	553.6	90.5	840	10	BF663213
36	553.4	90.5	725	9	AL527211
37	553.2	90.5	1122	10	BF663343
38	553.4	90.5	719	10	BG759156
39	553.2	90.5	760	10	B1114711
40	552.6	90.4	740	10	BG503437
41	551.8	90.3	988	10	B1520381
42	551.6	90.2	753	10	BF965976
43	551.4	90.2	1206	10	BM458518
44	551.2	90.2	733	10	BG496235
45	550.6	90.1	837	10	BF036445

ALIGNMENTS

RESULT 1

BG756624

LOCUS

DEFINITION 602713776F1 NIH_MGC_48 Homo sapiens CDNA clone IMAGE:4853825 5', 755 bp mRNA linear EST 15-MAY-2001

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Human
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgc.nci.nih.gov/.

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM1700 row: h column: 19

High quality sequence stop: 718.

Location/Qualifiers

1..755

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4853826"

/clone_lib="NIH_MGC_48"

/tissue_type="Primary B-cells from tonsils (cell line)"

/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library.

BASE COUNT 247 a 168 c 191 g 149 t

OPIGIN

Query Match 96.7%; Score 698.2; DB 10; Length 755;
Best Local Similarity 99.6%; Pred. No. 7e-138;
Matches 700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 12 gccaacatggtttcagcagcttcgtggaggttgccgggtggcctatgtctctcttggga 71
Db 2 GCCAACATCGGTTCICAGCGCTTCIGGAGGTTCGCCGGTGGCTATGCTCCTTGGGA 61

Oy 72 cctcatgccgaaattgctgcattgtaoatttattgatacagaagggctttgctc 131
Db 62 CTTCAATGCCGNAATTCGTCGATGTAGATGTTATGATCAACACAGCGCTTGGTC 121

132 gatgaaccttgactcaagtgagagacagaccatgctcttcaagtgcatacagctcact 191
122 GATGACCTTCACACTCAAGTCAGGAGACAGGCCATGCTTCATAGTCAGCTCAGCTAC 181

Oy 192 gatttcacctcaagttctcacagtgcccacagaaagtatctccacaaagctgcag 251
Db 182 GATTTTCCTCAAGTTTCCGCAGATGCCACCAGAGATGTTGTCACAAAGCTGCGAG 241

Oy 252 aaggcagacatcaatcacaaatggcgagcacacagatggccgaagaagattgaagccaga 311
Db 242 AAGGCAGACATCAATACAAATGGCGAGCCACACAGATGGGCCAASAGATTAAGGCCAGA 301

Oy 312 gaaggaaagccaagatacagatttgatgttttaadtattgaaggaaagaaatg 371
Db 302 GAAGGAAAGCCAAAGATGACAGATTTTGATCGTTTAAAGTTATGAAGGCAAGAAATG 361

Oy 372 agaacagaataatcaagaatgaagttaaagaactcaaaagacacactctctcgaagct 431
Db 362 AGGAACAGATAATCAGAAATGAAGTTAAGAGCTTCAAAAGGCGAGCTCTCTGAAAGCT 421

Oy 432 tctcccaaaaagcactctgactaaaggttactgctgctgctgctgctgctgctgctgct 491
Db 422 TCTCCCAAAAAGCACCCTGCTACTAAGGTACTGCTGCTGCTGCTGCTGCTGCTGCT 481

Oy 492 gct 551
Db 482 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541

552 ccagcccaagaggttctgcccgaagaagccacagggccagaaagcagcgcctgctccaaa 611
Db 542 CCAGCCCAAGAGGTTCCTGCCCCAAGAGCCACAGGCCAGAAAGCAAGCTGCTCCAAA 601

Oy 612 gctcagaaggtcgaagaagctccagccccaagaagaccctgctcgaagagcattcgaag 671
Db 502 GCTCAGAGGGTCMAAAGCTCCAGCCCAAGAAAGCAAGCTGCTGCTGCTGCTGCTGCT 661

Oy 672 aaagcataagtggaactcataaaaagtaataaaaggttctctttt 714
Db 662 AAAGCATAAGTGGCAATCATAAAGATTAATAGAGGTTCCTTTT 704

RESULT 2

LOCUS BG763626 822 bp mRNA linear EST 15-MAY-2001
DEFINITION 502735840f1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4861305 5'

ACCESION BG763626

VERSION BG763626

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 822)
NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rsb@nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1719 row: P column: 10
High quality sequence stop: 794.

FEATURES

source

Location/Qualifiers

1..822

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4861305"

/clone_lib="NIH_MGC_49"

/tissue_type="melanotic melanoma, high MDR (cell line)"

/lab_host="PH10B (phage-resistant)"

/note="Organ: skin; Vector: pOTR7; Site: 1: XhoI; Site: 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor: size

GGCAGCAG(G). Size-selected >500bp for average insert: size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH_MGC

Library. }

Library. }

BASE COUNT 250 a 179 c 206 g 177 t

ORIGIN

Query Match 95.0%; Score 693.2; DB 10; Length 822;
Best Local Similarity 99.2%; Pred. No. 8.1e-137;
Matches 708; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

Oy 9 gctccacacatggtttcagcagcttcgtggaggttgccgggtggcctatgtctctctt 58

Db 2 GCCGCCACATGGTGTTCAGGCCCTTCGTGGAGGTGGCCGGGTGGCTATGTCTCTCTT 61

Oy 69 ggaacctcatgcccggaaaaattgctgcgattgtagatgttattgatacaaacagagcttg 128

Db 62 GGACCTCATGCCGGAATAATGTCGCGATTGTAGATGTTATTGATCAGACAGGCTTIG 121

Oy 129 atcgatgaaccttgacactcaadtgaaggagacagagccatgcctttcaagtcagatcagctc 188

Db 122 GTCGATGAGCCCTTGCACCAAGTGGAGAGACAGGCCATGGCTTTTCAGTGCATGACGCTC 181

Oy 169 actgatttcactctcaagttctgcacagtgccaccagaaagtatgctccgaagcctag 248

Db 182 ACTGATTTTCCTCAAGTTTCGCGACAGTGCCACCACAGAGTATGTCGACAGGCTTGG 241

Oy 249 cagaagcagacatcaatacaaaaatggcagccacacacacatggcccaaaaatgaagccc 308

Db 242 CAGAAGGAGACATCAATACAAAATGGCGCCACACACCATGGCCAGAAAGATTGAAGCC 301

Oy 309 agagaaagaaagcccaadatatcacagattttgatcgttttaagttatgaagcagaagaaa 358

Db 302 AGAGAAAGGAAGCAAGATGACAGATTTTCATCGTTTAAAGTTATGAGGCCAAGAAA 361

Oy 369 atgaggaacaaataatacaagaatgaagtttaagaagcttcaaaagcagctctcctgaaa 428

Db 362 ATGAGGAACAGAAATTAATCAAGAAATGAAGTTAAGAAAGCTTCAAAAGGAGGCTCTCTCTGAAA 421

Oy 429 gcttctcccaaaaagcacccttgactaaaggtactactgctgctgctgctgctgctgct 488

Db 422 GCTTCTCCCAAAAGACCTGGTACTNANGGTA---CTGCTGCTGCTGCTGCTGCTGCT 478

Oy 499 gct 548


```
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: eye; Vector: pORF7; Site: 1; XhoI: Site_2;
EcoRI: cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(C). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library. [-
288 a 217 c 240 g 156 t

BASE COUNT 288 a 217 c 240 g 156 t
ORIGIN

Query Match 95.5%; Score 690.2; DB 10; Length 901;
Best Local Similarity 98.7%; Pred. No. 3.5e-136;
Matches 706; Conservative 0; Mismatches 8; Indels 1; Gaps 1:

Q: 9 gctgcacacatggtgttcaggcgcttcgtgaggttgccggatgagcctatgtctcttt 68
|||||
2 GCTGCCACATGTTTCAGCGCTTCGTGGAGTTGGCCGGTGGCCATGTCCTTT 61
|||||
59 ggaacctcgcgggaaattgctgcgattgagatgtattatcagaaacagggtttg 129
|||||
62 GGACCTCATGCCGAAATTCGTCGCAATGTAGATGTTATTCATCAGAACAGGGCTTG 121
|||||
129 gtcgatgaccttcactcaagtgaagaaagacagccatgcctttcaagtatcacagctc 189
|||||
122 GTCATGGACCTTTCACCTCAAGTGAGGAGACAGGCCAATGCTTTCAAGTGCATGCAGCTC 181
|||||
189 actgatttcctcctaagttcttcgacagtgccaccagaaagtatgtccacaaagcttg 248
|||||
182 ACTGATTTTCCTCAAGTTTCGCCACAGTGGCCACCAGAAAGTATGTCCGACAAAGCTTG 241
|||||
249 cagaagcagacatcaatacaaaatggcgagccacacagatggccaaagaagtgaagcc 308
|||||
242 CAGAAGCGACACATCAATACAAATGGCAGCCACACACGATGGCCCAAGAAAGATTGAAGCC 301
|||||
309 agagaaggaagcagaatgacagattttgattcgttttaaaattatgaaggcagaagaaa 358
|||||
302 AGAGAAAGGAAGCCCAAGATGACAGATTTTGATCGTTTAAAGTTTGAAGGCCAAGAAA 361
|||||
359 atgagaacaggaataatcaagaatgaagttaagaagcttcaaaaggcagctctcctgaaa 428
|||||
362 ATGAGGAACAGAAATATCAAGAAATGAAGTTAAGAAAGCTTCAAAAGGCAGCTCTCTTGAAA 421
|||||
429 gcttctcccaaaaagcaccctgactactaagggtactactactactactactactact 488
|||||
422 GCTTCTCCCAAAAAGACCTGTGTTACTAAGGGTACTGCTGCTGCTGCTGCTGCTGCT 481
|||||
489 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 548
|||||
482 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
|||||
549 gctccagccagaaaggttctgcccagaaagccacagggccagaaagcagc-gcctgctcc 607
|||||
542 GCTCCAGCCAGAAAGGTTCTGTCGCAAGAGCCACAGGCGCCAGAAAGCAGCGGCTGCTCC 601
|||||
608 aaaagctcagaagggtcaaaaagctccagcccgagaaagcaccctgcttcaaaagcattctg 657
|||||
602 AGAAGCTCAGAAAGGTCAAAAGCTCCAGCCCAAGAAAGCACTGCTCCAAAGGCATCTGG 651
|||||
658 cagaaagcagaagtggaacatcaaaaagtaataaaggttcttttgacctgtt 722
|||||
662 CAAGAAAGCATAAGTGGCAATCAACGAGTAATAGAGGTTCTTAATGAGCTGTT 716
|||||

RESULT 5
LOCUS BM083723
DEFINITION imagegc_2_2001/smk340bdf41.x1 NIH_MGC_16 Homo sapiens cDNA clone
IMAGE:4548493 5'. mRNA sequence.
ACCESSION BM083723
```

```
BM083723.1 GI:16951354
EST.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 797)
Kale,P.I., Harsch,I.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and
Prange,C.K.
The I.M.A.G.E. Consortium quality control effort: clone
resequencing for verification
Unpublished (2001)
Other_ESTs: BG474329
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been verified (found to hit its original self in the
correct orientation), as part of the I.M.A.G.E. Consortium quality
control effort. High quality sequence is defined as having 100 or
more base pairs with a phred quality value of 20 or greater, where
a sliding window of 4 base pairs with a phred quality value of 15
or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: LLCMI427 row: e column: 4
Seq primer: -21ml3
High quality sequence stop: 797.
FEATURES
Location/Qualifiers
1..797
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: eye; Vector: pORF7; Site: 1; XhoI: Site_2;
EcoRI: cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(C). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 241 a 174 c 200 g 176 t
ORIGIN

Query Match 95.3%; Score 688.4; DB 10; Length 797;
Best Local Similarity 99.4%; Pred. No. 8.4e-136;
Matches 700; Conservative 0; Mismatches 3; Indels 1; Gaps 1:

Q: 20 ggtgttcaggcgcttcgtgagagattgagccaggtgagcctatgtctcttggacctatgc 79
|||||
Db 1 GGTGTTTCAGCGCTTCGTGGAGGTGGCCGGTGGGCTATGCTCTCTTTGGACCTCATGC 60
|||||
80 cgaataattggtcgctgattatagatttatgatcagaaacagggctttgctcagtgacc 130
|||||
Db 61 CGGAAATTTGTCGCGATTGTAGATGTTATGATCAGAACAGGCTTTGGTCGATGACC 120
|||||
Q: 140 ttgactcaagttagagagacagccatgcctttcaagtgcatacagctactactat 199
|||||
Db 121 TTGACGTCGAAGTGAGGAGACAGGCCATGCTTTTCAAGTGCATGACGCTCACTGATTCA 180
|||||
Q: 200 cctcaagttctcacagtgccaccagaaagtatgccacaagcctgcagaaagcaga 259
|||||
Db 181 CTTCAAGTTTCCGACAGTGCCTCCACAGAAAGTATGTCGCAAGCCTGGCAGANGCAGA 240
|||||
Q: 260 catgaatacaaaatggcgagccacacagatggcccaagaadattgaagccagaaagaaa 319
|||||
Db 241 CATCAATACAAATGGCAGCCACAGCATGGGCCAAGAGATTGAAGCCAGACAAAGGAA 300
|||||
```

QY 320 agccaaatcacagatttttgcgttttaaaagtataaaggcaaaaataaaggaacag 379
 Db 301 AGCCAAATGACAGATTTTGAATGTTTAAAGTTATCAAGGCAAGAAATAGGAACAG 360
 QY 380 aataatcaaaatgaagttaagaagcttcaaaaggcagctctctctgaagcttctccaa 439
 Db 351 AATAATCAAGAAATGAAGTAAAGGCTTCAAGAGCAGCTCTCTGAAGTCTCTGCTG 420
 QY 440 aaagacccctggactaagggctactgctgctgctgctgctgctgctgctgctgctg 499
 Db 421 AAAAGCACCTTGGTACTAAGGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 QY 500 tgcgtgctaaagttcccaaaaagatcaccgcccgaagataaaagcctcccaaccca 559
 Db 481 TGTCTGCTGAAGTCTCCAGCAAAAAGATCACCGCGGAGTAAAGGTTCTAGCCCA 540
 QY 560 gaaggttccctgcccagaaagccagagccagaaagcagcctgctcccaaaagctcagaa 619
 Db 541 GAAGGTTCCTGCCAGAAAGCCACAGCCAGAAAGCAGCCGCTCTCTCAAGTACCTCAGAA 600
 QY 620 ggtcaaaaagctccagccagaaagccagcctgctcccaagggctctgagcaaaagcata 679
 Db 601 CGGTCAAAAAGCTCCAGTCCAGAAAGCAGCTGCTCCAAAGGCTCTGCGCAAAAGGATA 660
 QY 680 agtggcaatcataaaagtaataaa-gattctttttgacctgatt 722
 Db 661 AGTGGCANTATATAAAGTAAATNAAGGGTCTCTTTTTCACCGTGT 704

RESULT 5
 AUI23120 728 bp mRNA line: EST 23-OCT-2000
 LOCUS AUI23120 NT2RM1 Homo sapiens cDNA clone NT2RM1001-50 5', mRNA
 DEFINITION sequence.
 ACCESSION AUI23120
 VERSION AUI23120.1 GI:10947836
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 728)
 Ota.T., Nishikawa,T., Suzuki,Y., Kawai,Y., Ishii,S., Saito,K.,
 Nakamura,Y., Nagai,I., Sugano,S. and Isogai,T., Suzuki,Y., Kawai,Y.,
 HPI human cDNA project (Ota,I., Nishikawa,T., Suzuki,Y., Kawai,Y.,
 Ishii,S., Saito,K., Nakamura,Y., Nagai,I., Sugano,S., Isogai,T.)
 Unpublished (2000)
 CONTACT Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HPI human cDNA project; 5', 3'-end one pass sequencing; Helix
 Research Institute; cDNA library construction; Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 Location/Qualifiers
 1..728
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RM1000750"
 /clone_lib="NT2RM1"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /note="Vector: pUC19FL3; mRNA from undifferentiated NT2 neuronal
 precursor cells"

BASE COUNT 220 a 173 c 181 g 153 t 1 other:
 ORIGIN

FEATURES

source
 1..728
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RM1000750"
 /clone_lib="NT2RM1"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /note="Vector: pUC19FL3; mRNA from undifferentiated NT2 neuronal
 precursor cells"
 220 a 173 c 181 g 153 t 1 other:
 Query Match 94.3% Score 690.6 DB 5; Length 728;

Best Local Similarity 98.9%; Pred. No. 3.7e-134;
 Matches 715; Conservative 0; Mismatches 5; Indels 3; Gaps 3;
 QY 1 cgcctaacgctgcacaatggtttcagcgcttcgtggaggttgccgggtgctctatg 61
 Db 5 CCGCTAACCTGCCAATGGTGTTCAGGGCTTCGTGGAGGTGGCCGAGTGGCCTATG 65
 QY 51 tctcctttggacctcctgcggaaaaattggctcgcgattgtagatgtattgatacaaca 120
 Db 66 TCTCTTTGGACCTCAATGCGGAAATTTGGTCGCGATTGTAGATGTATTGATCAGAACA 125
 QY 121 ggcctttgctcaatgaaccttgcactcaagtggagagacagggccatgcctttcaaatgca 140
 Db 126 GGCCTTTGGTGAATGACCTTGCATCAAGTAGGAGACAGGCCATGCCCTTCAGATGCA 135
 QY 181 tgcagctcactgatttctcctcctcaagtttctgcacagtgcccccacaaagtatgtccgac 210
 Db 186 TGCAGCTCACTGATTTCCTCTCAAGTTTCCGCACAGTGGCCACCAAGATATGTCCGAC 215
 QY 241 aaccttggcagaaggcagacatcaatacaaaatggcagccacacatggagcccaaga 340
 Db 245 AAGCCTGGCAAGAGGAGACATCAATACAAATGGCAGCACACGATGGGCCAAGAAGA 335
 QY 301 ttaaaaccagagaaagaaagccaaagatgacagattttgatcttttaaaagttatgaag 360
 Db 306 TTGAAGCCAGAGAAAGAAAGCCAAAGATGACAGATTTTGATCGTTTAAAGTTATGNAG 355
 QY 361 caaagaaatggagaaacagagataatcaagaatgaagttaagaagcttcaaaaggcagctc 420
 Db 366 CAAAGAAAGTGAGGAACAGAGAAATATCAAGATGAAGTTAAGAGCTTCAAAAGGCAGCTC 415
 QY 421 tctgaaagcttctcccaaaaagccctgactaagggtactgctgctgctgctgctg 430
 Db 426 TCTGAAAGCTTCTCCCAAAAAGCACCCTGCTACTAAGGGTACTGCTGCTGCTGCTG 415
 QY 481 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
 Db 486 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515
 QY 541 gtaaaagctccagccagaaaggttctctgcccagaagccagccagccagaaagcagc 610
 Db 545 GTAAAAGGCTCCAGCCAGAGAGGTTCTCTGCCAGAAAGCCACAGCCGCAANAAAGCAGCGC 605
 QY 601 ctgctccaaaagctcagaaggttcaaaaagctccagcccccagagacactgctccaaagg 640
 Db 606 CTGCTCCAAAAGCTCAGAAGGGTCAAAAAGCTCCAGCCCAAGAGCA-CIGCTCCAAAGG 634
 QY 661 catctggcagaagaagcataag-tggcaatcat-aaaaagtaataaaggttcttttggacc 718
 Db 665 CATCTGCGAAGAAAGCATATAGTTGGCNAATCATANAAAGTAATANAGGTTCCTTTTGAAC 724
 QY 719 tgtt 722
 Db 725 TGT 728
 RESULT 7
 AUI231631 818 bp mRNA line: EST 13-FEB-2001
 LOCUS AUI231631 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM002YH21 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AUI231631
 VERSION AUI231631.1 GI:12795124
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 818)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 CONTACT Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Location/Qualifiers
1..818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS00M002YH21"
/clone_lib="L11_NFL001_HBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 5; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 5 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 510 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"

FEATURES
source
BASE COUNT 271 a 172 c 199 g 175 t 1 others
ORIGIN

Query Match 94.2%; Score 680.4; DB 9; Length 818;
Best Local Similarity 99.0%; Pred. No. 4.1e-134;
Matches 697; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

Qy 19 ttgtgttcaggcgttcgtggagggttgccgggtgagccctatctctcttttgacacctatg 78
Db 1 TGGTGTTCAGGCCTTCGTGGAGGTGGCGGGTGGCCATGTCCTTTTGACCTCATG 50

Qy 79 ccggaataatgttcgcgatttagatgttatgacagacagggtttgtccatggaac 138
Db 61 CCGGAAATATGTCGCGATTGTAGATGTTATGATCAGAACAGGCGTTTGGTCGATGGAC 120

Qy 139 cttgcactcaagtgaagagacaggccatcccttcaagtgcagctcactgatttca 198
Db 121 CTTGCACTCAAGTAGAGACAGGCCATGCCCTTCAAGTGCATGACGTCATGATCA 180

Qy 199 tctcaagttctgtacatgcccaccagaagtatgtccgacaagcctggcagaaggag 258
Db 181 TCTCAAGTTTCGCGACAGTGGCCACCAGGAAGTATGTCGACAGCGCTGGCAGAGGCGAG 240

Qy 259 acatcaatacaaaatgggcagccacacgctggcccaaaagatgaagccagaagaaga 318
Db 241 ACATCAATACAAAATGGGAGCCACACAGATGGCCACAGAGATTTGAAGCCAGAGAAAGA 300

Qy 319 aagccaagatgacagattttgatcggttttaagttatgaagcaagaagaaatgaggaac 378
Db 301 AAGCCAAGATGACAGATTTTGTATCGTTTAAAGTTATGAAGCAAGAAATGAGCAACA 360

Qy 379 gaataatcaagaatgaagtaagaagcttcaaaaggcagctctctgaaagcttctcca 438
Db 361 GAATAATCAAGAAATGAAGTTAGAAGCTTCAAAAGCGAGCTCTCTGAAAGCTTCTCCA 420

Qy 439 aaaaacacctgtactaaggtaactgctgctgctgctgctgctgctgctgctgctgctg 498
Db 421 AAAAAGCACCTGGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 474

Qy 499 ctgctgctgctaaagttccagcaaaaagatcacccgcgcgcgcgcgcgcgcgcgcgcgc 558
Db 475 CTGCTGCTGCTAAAGTTCACGCAAAAAGATCACCCGCCGCGAGTAAAAGGCTTCAGGCC 534

Qy 559 agaaggttctctccagagaagccagagccagaagaagcagcgcctgctccaaaagctcaga 618
Db 535 AGAAGGTTCTTCGCCAGAAAGCCACAGGCCAGAAAAGCAGCGCTTCTCAAAAAGCTCAGA 594

Qy 619 aggggtcaaaaagctccagcccgagagacacctgctccaaagagcatctggcagaagaagcat 678
Db 595 AGGGTCAAAAAGCTCCAGCCCGAGAAAGACCTTCTCCAAAGGCGATCTGGCAAGAAAGCAT 654

Qy 679 aagtgccaatcataaaaaagtaataaaggtttctttttgacctatt 722
Db 655 AAGTGGCAATCAATAAAAGTATAAAGGTTCCTTTTGTACCTGTT 698

RESULT 8
BM472466 723 bp mRNA linear EST 05-FEB-2002
LOCUS AGENCOURT_6464185 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:553444
DEFINITION 5', mRNA sequence.
ACCESSION BM472466
VERSION BM472466.1 GI:18521508
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 723)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DIP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1231 row: f column: 13
High quality sequence stop: 637.
Location/Qualifiers
1..723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5538444"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 235 a 167 c 178 g 143 t
ORIGIN

Query Match 94.2%; Score 679.8; DB 10; Length 723;
Best Local Similarity 98.0%; Pred. No. 5.5e-134;
Matches 703; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Qy 5 aacgtcccaacatagtttcaaggccttcgtgaaggttgccgggtgacctatcttcc 65
Db 1 AACGCCCAACATGGTGTTCAGGCGCTTCGTGGAGGTGGCCGGTGCCCTATGTCTCC 50

Qy 66 ttggacctcatgccggaaaattgctgcgattgtagatgttattgatcagaacaggct 125
Db 61 ITTGACCTCATGCCGGAATTTGGTCGGATTGTAGATTATTGATCAGAACAGGCT 120

Qy 126 ttggtcgatgaccttgcaactcaagtggaggagacaggccatgccttccaagtgcacg 185
Db 121 TTGTGCGATGGACCTTGCACTCAAGTGAGGAGACAGGCGCATGCCCTTCAAGTGCGACG 180

Qy 186 ctcactgatttcactcctcaagtttctgcacagtgccaccagagaagtatgccgacaagcc 245
Db 181 CTCACGTATTTCATCCTCAAGTTTCCGACAGTCAGTGGCCACCAAGATATGTCGACAAGCC 240

Qy 246 tggcagaaggcagacatcaatacaaaaatgggcagccacacagatggcccaagaagattgaa 305
Db 241 TGGCAGAGGCGACAGATCAATACAAAATGGGCGAGCCACACGATGGCCCAAGAAATTGNA 300

```

OY 305 gccagagaaagaaacccaagatgacagattttgatcgttttaaaagttaagagagaaag 365
Db 301 GCCAGAGAAAGAAAGCAAGATGACAGATTTTGAATGTTTTAAAGTTATCAAGCCAAAG 360
OY 365 aaataggaacagaataatcaagaatgaagttaagaagttcaaaagggcagctctcctg 425
Db 361 AAAATGAGGAACAGATAATCAAGATGAATTAAGAGGTTTCAAAAGGCAAGCTCCTG 420
OY 425 aaagcttctcccaaaagacacctgatactactaagggtactgctgctgctgctgct 485
Db 421 AAAGCTTCTCCCAAAAGACCTGGTACTTAAGGGTA-----CTGCTGTGCT 458
OY 485 gctgctactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 545
Db 469 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
OY 545 aaggctccagccagagaggttctgctcccaaaagccacagggccagagaaagcaagctgct 605
Db 529 AAGGCTCCAGCCAGAGGTTCTGCCAGAAAGCCACAGCCAGAAAGCAAGCTGCTGCT 588
OY 605 ccaaaagctcagaaggtcaaaagctccagccagagagcactgctccaaagggatct 665
Db 589 CCAAAAGCTCAGAGGGTCAAAAGGTTCCAGGTCAGGTCAGAAAGCAAGCTGCTGCTGCT 648
OY 665 ggcagaaagacataagtggaatcataaaagtaataaaggttcttttgaactctt 722
Db 649 GGCAGAAAGCATAGTGGCAATCATAAAAGTAATAAGGTTCTTTTGAATGCTGTT 705

RESULT 9
BG395368 730 bp mRNA line: EST 12-MAR-2001
LOCUS 502459230F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581479 5'
DEFINITION mRNA sequence.
ACCESSION BG395368
VERSION BG395368.1 GI:13289914
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Euthera: Primates: Catarrhini: Hominiidae: Homo.
TITLE NIH-MGC http://imgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: AICC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: http://imgc.llnl.gov
Plate: L12W1302 row: d column: 24
High quality sequence stop: 727.
Location/Qualifiers
1..730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4581479"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: eye; Vector: pCIB7; Site_1: XhoI; Site_2:
EcoRI; Organ made by: oligo-dT priming. directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACAGAG(G). Library constructed by Ling Hong
in the Laboratory of Gerald W. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
231..a 170..c 183..g 145..t

```

BASE COUNT
CYPIN

Query Match 94.0%; Score 679; DB 10; Length 730;
Best Local Similarity 99.0%; Pred. No. 8.le-134;
Matches 704; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```

OY 12 gccacaatgatgtttcaggcgcttcgtagaggttgccgggtggcctatgctctcttggga 71
Db 1 GCCAACATGGTGTTCAGGCGCTTCGTGGAGGTTGGCCGGGTGCTATGCTCTCTTTGGA 60
OY 72 cctcatcccgaaaaatggtcgcgattgtagatgttatgatcagaacagggcttggctc 131
Db 61 CTTATGTCGGGAAATTTGGTCGGGATTTAGATGTTATTGATCAGAACAGGGCTTTGGTC 120
OY 132 gatggacttgcactcaagtggagagacagggccttctcaagtcatgcatgcatcact 191
Db 121 GATGGACCTTGCACATCAATGAGGAGACAGCCATGCTTTTAAAGTGCATGCAGCTCACT 180
OY 192 gattcatctcgaagtttctgcacagtgccaccagaaagtatgtccaaagcctggcag 251
Db 181 GATTTCATCTCAAGTTTCCGCACAGTCCACAGAGTATGTCCAGACAGGCTTGGCAG 240
OY 252 aaggcagacatcatatacaaaatgggcagccacacgattggccagaaagattgaagccaga 311
Db 241 AAGGCACACATCAATACAAAATGGGCAGCCACAGATGGGCCAGAGATTTGAAGCCAGA 300
OY 312 gaaagggaaagccaaagatgcagattttgatcgtttttaaagtattgaaggcaaaagaaatg 371
Db 301 GAAAGGAAGCCAGATGACAGATTTTGATCGTTTTTAAAGTTATGAAGGCAAGAAAGATG 360
OY 372 agaaacagaataatcaagaatgaagtttaagaagctcaaaagcagctctcctaaagct 431
Db 361 AGGAACAGAAATATCAAGAATGAAGTTAAGAAGCTTTCAAAAGGCGCTCTCTTGAAAGCT 420
OY 432 tctcccaaaagccactgctactaagggtactgctgctgctgctgctgctgctgctgct 499
Db 421 ICTCCCAAAAAGCACCTGGTACTAGGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
OY 492 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 551
Db 481 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
OY 552 ccagccagagaagttctctgcccagaagccacagggccagaaagcagcgcctgctccaaaa 611
Db 541 CCAGCCCAAGAGGTTCTGCCCAAGAACCCACAGGCCAGAAAGCAGCGC-CTGTCCAAAA 599
OY 612 gctcagaagggtcaaaaagctccagcccaagaagcactgctccaaaggcatctggcaag 671
Db 602 GCTCAGAAAGGTCAAAAGGCTCCAGCCAGAAAGACCTGCTCCAGAGGCACTTGGCAAG 659
OY 672 aaagcataagtcgaatcataaaagtataaagggttcttttgaaccttt 722
Db 660 AAAGCATAAGTGGC-ATCATATAAAGTAATAGAGGTTCTTTTTCGACCTGGT 709

```

RESULT 10
BI914072

```

LOCUS 603182003F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246184 5'
DEFINITION mRNA sequence.
ACCESSION BI914072
VERSION BI914072.1 GI:16178382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS Mammalia: Euthera: Primates: Catarrhini: Hominiidae: Homo.
TITLE NIH-MGC http://imgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov

```

Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11621 row: e column: 01
 High quality sequence stop: 713.
 Location/Qualifiers
 1..746
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="5246184"
 /clone_lib="NIH_MGC_121"
 /lab_host="DH10B"
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 21 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb. Insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."
 235 a 175 c 184 g 151 t

BASE COUNT
 ORIGIN

Query Match 92.8%; Score 677.4; DB 10; Length 746;
 Best Local Similarity 99.0%; Pred. No. 1.8e-133;
 Matches 592; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 17 catgtgttcacagccttgtagaggttggccgggtgacctatgtctctttgacctca 75
 DB 1 CAATGGTGTACGGCTTCGTGGAGGTTGGCGGGTGGCTATGCTCTCTTGGACCTCA 60

OY 77 tcccgaaaattggtcgcatgttagatgttattcatcagaacaggctttggtcgatcg 135
 DB 61 TCCCGAAAATTGGTCGGGATTGTAGATGTATTATCATCAGACAGGGCTTTGGTCAATG 120

OY 137 accttgcactcaagttagaggaacaggccatgcctttcaagtgcagcctcactgattt 195
 DB 121 ACCTTTGCACCTCAAGTGAGGACAGAGCCATGCTTTTCAAGTGCATGCACCTCATTT 180

OY 197 catctcaagatttctgcacagtgtccacacagaagatgtatcgcagaagcctgcacaaaggc 256
 DB 181 CATCTCAAGTTTCCGCACAGTTCGCCACACAGAAATATGCCACAAAGGCTGGCAAGG 240

OY 257 agacatcaatacaaaatggcagccacacacgatggcccaagaagattgaagcagagaag 316
 DB 241 AGACATCAATACAAATGGCGACCCACACGATGGGCCAAGAGATTGAAGCCAGTGAAG 300

OY 317 gaaagccaaagtacagatatttgatgttttaagtattgaaggcagaagaaatagaga 375
 DB 301 GAAAGCCAAAGATGACAGATTTTGTATGCTTTTAAAGTTATGAAGGCCAAGAAAATGAGGAA 360

OY 377 cagaataatcaagaatgaagttaagaacttcaagaagcagactctctgaaagccttccc 436
 DB 361 CAGAATAATCAAGAATGAAGTTAAGAAGCTTCAAAAGGCAGCTCTCTCTCAAGCTTCTCC 420

OY 437 caaaaaagcacccttgactaagggttactgtctgtctgtctgtctgtctgtctgtctgtc 496
 DB 421 CAAAAAGCACCCTGTACTAAGGTTACTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 480

OY 497 tctctactgtctaaagttccagcaaaaaaatcatccgcgcgcgagtagtaaaagcctccag 556
 DB 481 TCTCTCTGTCTAAAGTTCCAGCAAAAAAAGATCATCCGCCCGGAGTAAAGGCTCCAGC 540

OY 557 cagaagattctctccagaaagccacagagccagaaagacagcctgtctcaaaaagctca 616
 DB 541 CCAGAGAGGTTCTTCCCAAGAAAGCCACAGGCCAGAGAGAGCGGCTGTCTCAAGAGCTCA 600

OY 517 caagggtcaaaaagct-ccagcccaagaagaccctgctctccaaagcattctgacagaaga 675
 DB 601 GAAGGGTCAAAAAGTCCCAAGAGACACCTGCTCTCAAGAGCATCTGCGCAAGAAAG 660

OY 675 cataagtgacatcataaaaagataataaaggttcttttt 714
 DB 661 CATAGTGGCAATCATACCAAGTATATAAGGTTCITTTT 699

RESULT 11
 BG260421 732 bp mRNA linear EST 13-FEB-2001
 LOCUS 602371552F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4479420 5'
 DEFINITION mRNA sequence.
 ACCESSION BG260421
 VERSION BG260421.1 GI:12770237
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 732)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: csapbs@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10311 row: h column: 13
 High quality sequence stop: 718.

FEATURES
 source

1..732
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4479420"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 225 a 175 c 185 g 147 t

Query Match 93.5%; Score 675.2; DB 10; Length 732;
 Best Local Similarity 98.6%; Pred. No. 5.1e-133;
 Matches 702; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

OY 2 gcctaacactgcaacatggttctcagcgtcttcgaaggttgccgggtgacctatgt 61
 DB 1 GCCTAACCTGCAACATGGTGTTCAGGCGCTTCGTGGAGGTGGCGGGTGGCCTAIGT 60

OY 62 ctctcttgagacctatgccgaaaaattggtcgcatgttagatgttattgatcagaacag 121
 DB 61 CTCCTTTGGACCTCATSCGGGAAATTTGTCGCGATTGTAGATGTTATTGATCAGAACAG 120

OY 122 ggccttgatgatgacaccttgacctcaagttaggagacaggccatgcctttcaatgcat 181
 DB 121 GCGTTTGGTTCGATGGACCTTGCACCTCAAGTGAAGAGACAGGCCATGCCCTTTCAAGTGCAT 180

OY 182 ccagctcactaatttcactcctcaagtcttcacagatccaccacagaagatgtctcgaca 241
 DB 181 GCAGCTCACTGATTTTCACTTCCTCAAGTTTCCGACAGTTCGCCACCAAGAGTATGTCGACA 240

242 agctgacgagagcagacacatcaatacaaaaatggcgagccacacacatggcgcaagaagat 301
 Db AGCTGGCAGAAAGCAGACATCAATACAAAATGGCAGCCACACGATGGGCAAGAAGAT 300
 Oy tgaagccagagaaaggaagcagaatgacacatgttgatcgttttaagttatgaagc 361
 Db TGAAGCCAGAGAAAGGAAGCCAGATGACAGATTTTGTGCTTTTAAAGTTATGAAGGC 360
 Oy aagagaaatgaggacagaataatcaagaatgaagtttaagaagcttcaaaagcagctct 421
 Db AAAGAAAATGAGGAACAGATAATCAAGAATGAAGTTAAGAAGCTTCAAAAGCCAGCTCT 420
 Oy cctaaagcttctcccaaaaagcaccctgtactactaaaggtactcctgtctcgtcgc 481
 Db CCTGAAAGCTTCTCCCAAAAAGCACCCTGGTACTTAAGGGTACTGCTCTGCTGCTGC 480
 Oy tctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 541
 Db TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Oy taagaagcttccagcagaaggttctgcccgaagacacacacacacacacacacacacac 601
 Db TAAAGAGCTCCAGCCAGAAAGGTTCTGCTCCAGAAAGCCACAGGCGGAGAGCAGAGGC 600
 Oy tctcccaaaagctcagaaggtc-aaaaagctccagcccaaaaagcagcagcagcagcagc 660
 Db TGTCCAAAAGCTCAGAAGGTCNAAAAAGCTCCAGCCCAAGAAGCACTGCTCCAGAGG 660
 Oy catctgcagaagaagcagaagtcacatcaaaaagtaataaaggtctttt 712
 Db CATCTGCAAGAAAGCAATAGTGGC-ATCATAAGAAGTACTAAGGTTCCTTT 711

RESULT 12

BM043710
 LOCUS 50352050ZF1 NIH_MGC_40 705 bp mRNA linear EST 07-NOV-2001
 DEFINITION mRNA sequence.
 ACCESSION BM043710
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rstraubs@mail.nih.gov
 Tissue Procurement: DCID/DTP
 CDNA Library Preparation: Ling Hong/Pubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov/
 Plate: LHC1928 row: c column: 01
 High quality sequence stop: 704.
 Location/Qualifiers
 1..705
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:546200"
 /clone_lib="NIH_MGC_40"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: prostate; Vector: pOTF"; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by

FEATURES
Source

RESULT 13
 AL528184
 LOCUS AL528184 LTI_NFL003_NBC3 712 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL528184 prime, mRNA sequence.
 ACCESSION AL528184
 VERSION AL528184.1 GI:12791677
 KEYWORDS EST
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: This is a NIH_MGC Library.
 BASE COUNT 215 a 165 c 182 g 142 t
 ORIGIN

Query Match 93.4%; Score 674.4; DB 10; Length 705;
 Best local Similarity 98.7%; Pred. No. 7.6e-133;
 Matches 591; Conservative 0; Mismatches 6; Indels 3; Gaps 1;
 Oy 25 caggccgttcgtgaggttgccgggtggcctatgtctcttttggaccctatgccgaaa 85
 Db caggccgttcgtgaggttgccgggtggcctatgtctcttttggaccctatgccgaaa 61
 Oy 85 attgctcgagattgtaagtattatcacaagaaggtttgtcgatgagccttcac 145
 Db attgctcgagattgtaagtattatcacaagaaggtttgtcgatgagccttcac 121
 Oy 145 tcaagtgaagagagagccatgccttttcaagtgcacgtcactgatttcctcctca 205
 Db tcaagtgaagagagagccatgccttttcaagtgcacgtcactgatttcctcctca 181
 Oy 205 attcttcacagtcgccacacagaagtatgtccaaagcctggcagaagcagacatcaa 265
 Db attcttcacagtcgccacacagaagtatgtccaaagcctggcagaagcagacatcaa 241
 Oy 265 tacaataatggcagccacacacatgggccagaagattgaagccagagaaggaagccaa 325
 Db tacaataatggcagccacacacatgggccagaagattgaagccagagaaggaagccaa 301
 Oy 325 gatcagagattttgatcgttttaagtattgaaggcaagaataatgagaacagataat 385
 Db gatcagagattttgatcgttttaagtattgaaggcaagaataatgagaacagataat 361
 Oy 385 caagaatgaagttaagaagcttcaaaagcagctctcctgaaagcttctcccaaaaagc 445
 Db caagaatgaagttaagaagcttcaaaagcagctctcctgaaagcttctcccaaaaagc 421
 Oy 445 acctgtactaagggtta---ctgctgctgctgctgctgctgctgctgctgctgctg 502
 Db acctgtactaagggtta---ctgctgctgctgctgctgctgctgctgctgctgctg 481
 Oy 502 tgcgtctaaagtccagcaaaaagatcaccccgcgagtataaaagcttccagccagaa 562
 Db tgcgtctaaagtccagcaaaaagatcaccccgcgagtataaaagcttccagccagaa 541
 Oy 562 ggttcctgccagaagaagccacagccagaaagcagcgcctgtccaaaagctcagaaggg 622
 Db ggttcctgccagaagaagccacagccagaaagcagcgcctgtccaaaagctcagaaggg 601
 Oy 622 tcaaaagctcagccagccagaaagcagccttccaaagcctctggcagaagcataagct 682
 Db tcaaaagctcagccagccagaaagcagccttccaaagcctctggcagaagcataagct 661
 Oy 682 ggcacatcaaaaagtaataaaggttttttgacctgtt 722
 Db ggcacatcaaaaagtaataaaggttttttgacctgtt 701

REFERENCE 1 (bases 1 to 712)
Li, W.B., Gruber, C., Jesse, J. and Polayres, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91005 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0PC025YL02"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 5; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five primers
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 5
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com UBL :
http://fulllength.invitrogen.com"

BASE COUNT 227 a 161 c 177 g 145 t 2 others
ORIGIN

Query Match 93.3%; Score 673.4; DB 9; Length 712;
Best Local Similarity 98.3%; Pred. No. 1.2e-132;
Matches 593; Conservative 1; Mismatches 12; Indels 9; Gaps 1;

Oy 18 atggttgcaggccttcgtgaggttgccgggtgagcctatctcccttgacctat 77
Db 1 ATGGTGTTGAGGCGCTTCGTGGAGTTGGCCGGTGGCTATGTCCTTTGACCTCAT 50
Oy 78 gccggaattggtcgcgatttagatgtattgatcagaacaggccttgatcgtgga 137
Db 61 GCCGGAAATTTGTCGCGATTGTAGATGTTATGTATCAGAACAGGCGCTTGGTCGATGCA 120
Oy 138 ctttgactcaatgaggagacagggccttccttcaagtgcagctcactcaatttc 197
Db 121 CTTTGCACTCAAGTGAAGAGACAGGCGCATGCCCTTCAAGTGCATGCAGCTCACTGATTTTC 180
Oy 198 atcctcaagtttgcacagtccaccacagaagatgtgtccgacaagcctggcagaaggca 257
Db 181 ATCCCTCAAGTTTCCGCACAGTGCACACAGAGATGTGTTCGACAAAGCCCTGGCAGAGGCA 240
Oy 258 gacatcaatacaaaatgggcagccacacagatggcccaaaagattgaagccagagaaagg 317
Db 241 GACATCAATACAAATGGCCACCCACGATGGCCCAAGAAATTAAGCCAGAGAAAGG 300
Oy 318 aaagccaagatacagattttatcgttttaagttatgaagcagaagaaatagggaac 377
Db 301 AAAGCCCAAGATGACAGATTTTATGCTGTTTAAAGTTATGAAGGCAAGAAATGAGGAAC 360
Oy 378 agataatcaagaataaagttaagaagcttcaaaagcagcttctctaaagcttctccc 437
Db 361 AGAATATCAAGAAATGAAGTTAAGAGCTTCAAAAGGCGAGCTCTCTGAAAGCTTCTCCC 420
Oy 438 aaaaagccacttgactaagggtagctgctgctgctgctgctgctgctgctgctgctgct 497
Db 421 AAAAAAGCACTGGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
Oy 498 gctgctgctgctlaaagttccacaaaagatcacccgcgcagagtaaaaggctccagcc 557
Db 472 GCTGTGCTGCTTAAAGTTCACGCAAAAAGATCACCGCGCGAGTTAAAGGCTCCAGCC 531
Oy 558 cagaaggttccctgcccaagaaagccagagccagaaagcagccctgctcccaaaagctcag 617
Db 558 cagaaggttccctgcccaagaaagccagagccagaaagctcagccctgctcccaaaagctcag 617

Db 532 CAGAAGTTCTCTCCAGAAAGCCACAGCCAGCAAGAGCGGCTCTCTCCAAAGCTCAG 591
Oy 618 aagggtcaaaaagctccagccagaaagcactgctcccaagcatctgctgcaagaagca 677
Db 592 AAGGGTCAAAAAGCTCCAGCCAGAAAGCACCTGCTCCAAAGGCACTCTGGCAAGAAGCA 651
Oy 578 taagtgcgaatcataaaaagtaataaaggcttctttttgacctgtt 722
Db 552 TAGTGGCAATCAATAAAGTAATAAAGGTTCTTTTGGACCTGT 696

RESULI 14
BM480399
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM480399 1116 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6458573 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5574262
5', mRNA sequence.
BM480399
BM480399.1 GI:18529441
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1116)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cchapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1232 row: j column: 23
High quality sequence start: 27
High quality sequence stop: 636.
Location/Qualifiers
1. 1116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5574262"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH-MGC Library."

BASE COUNT 298 a 308 c 254 g 211 t 45 others
ORIGIN

Query Match 93.0%; Score 671.4; DB 10; Length 1116;
Best Local Similarity 98.2%; Pred. No. 3.5e-132;
Matches 594; Conservative 0; Mismatches 1; Indels 12; Gaps 1;

Oy 15 acatggttccagcgttctgaggttgccgggtgagcctatgctcctttgacctc 75
Db 38 ACATGGTGTTCAGGCGCTTCGTGGAGTTGGCCGGTGGCTATGCTCTCTTGGACCTC 97
Oy 75 atgccgaaaaattggtcgcgatttagatgtattgatcagaacaggcttggatcag 135
Db 98 ATGCCGAAAATTTGTCGCGATTGTAGATGTTATTGATCAGAACAGGCTTTGGTCGATG 157
Oy 175 gaccttgacctcaagtgaaggagcagccatgcctttcaaatgcatgcagctcactgatt 195
Db 158 GACCTTGCACTCAAGTAGAGGACAGAGCCATGCCCTTTCAAGTGCATGCAGCTCACTGATT 217
Oy 195 tcactctcaagtcttcacagagtcacacagagtcacacagagtcacacagagtcacagag 255

```
|||||
1b 218 TCACTCTCAAGTTCGGACAGTGCACCAAGATGTCGCAAGACCTGGCAGAAAG 277
QY 256 cagacatcaatacaaaatgggcagccacacagatggcccaagaagattcaagccagagaa 315
Db 278 CAGACATCAATACAAATGGGAGCCACAGGATGGCCCAAGAGATTGAAGCCAGAGAA 337
QY 316 ggaagcccaagatgacagatttttgatcgttttaagattatgaagcgaagaaatgagga 375
Db 338 GGAAGCCCAAGAAGACAGATTITGATCGTTTAAAGTTAAGGCAAGAAATGAGGA 397
QY 376 ecagaataatcaagaataaagttgaagcttcaaaagcagctctctgaaagctttctc 435
Db 398 ACAGAAATCAAGATCAAGTTAGAGCTTCAAAAGGCAAGCTCTCTGAAAGCTTCTC 457
QY 436 ccaaaaaagccctgtactaaggttactgctgctgctgctgctgctgctgctg 495
Db 458 CCAAAAAAGCCCTGGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCTG 505
QY 496 ctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 555
Db 506 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565
QY 556 cccagaaaggttctgcccagaaagccacagccagaaagcagcagcagcagcagcagc 615
Db 566 CCCAGAAAGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625
QY 616 gaaaggttcaaaagctcagccagaaagcagcagcagcagcagcagcagcagcagc 675
Db 626 AGAAGGGTCAAAAGCTCCAGCCAGAAAGCAGCTGCTGCTGCTGCTGCTGCTGCTG 685
QY 676 cataagtgcccaatcetaaaagtaaaagattcttttaacctgt 722
Db 686 CATAAGTGGCAATCATAAAGATTAATGAAGGTTCTTTTGAACCTGTT 732

RESULT 15
LOCUS BG437749 724 bp mRNA linear EST 14-MAR-2001
DEFINITION 602489826F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4521783 5',
RNA sequence.
ACCESSION BG437749
VERSION BG437749.1 GI:13344255
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgs@fmail.nih.gov
Tissue Procurement: DCID/DIF/Gazdar
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1385 row: d column: 08
High quality sequence stop: 704.
Location/Qualifiers
1..724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4521783"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCM7; Site: 1; XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
```

FEATURES

SOURCE

Search completed: June 21, 2002, 13:49:45
Job time: 5885 sec

into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 229 a 165 c 185 g 144 t
ORIGIN

Query Match 92.8%; Score 670; DB 10; Length 724;
Best Local Similarity 98.3%; Pred. No. 6.5e-132;
Matches 700; Conservative 0; Mismatches 5; Indels 7; Gaps 2;

QY 12 gccaacatggtgttcagagccttcgtgaggttgccgaggtggcctatgtctcttggga 71
Db 1 GCCAACATGTTTCAGGCGCTTCGTGAGGTTGGCCGGTGGCTATGTCTCTTTGGA 60
QY 72 cctcatgccggaaaaattggtcgcgattgtagattattgatcagaaacagagctttggtc 131
Db 51 CCTCATGCCGGAAAAATGGTCGCAITGTAGATGTTATTGATCAGAACAGGGCTTTGGTC 120
QY 132 gatgagccttcacacaaagtgagagagcagccatgctttcaagtcacatgcagctcaat 191
Db 121 GATGGACCTTGCACCTCAAGTGAAGGAGACAGGCCATGCTTTCAAGTGCATGCAGCTCACT 180
QY 192 gatttcattctcaagtttctcagagtcgacagtcgacagtcgacagtcgacagtcgacag 251
Db 181 GATTTCATCTCANTTTCCGACAGTGGCCACCAAGATGATGTCCGCAAGCTTGGCAG 240
QY 252 aagcagacatcaatacaaaatgggcagccacacagtcgacagtcgacagtcgacagtcgacag 311
Db 241 AAGCAGACATCAATACAAATGGCAGCCACAGCAGTGGCCAGAAAGATTGAAGCCAGA 300
QY 312 gaaaggaagcccaagatgacagattttgattgcttttaagttatgaaggcaagaaatg 371
Db 301 GAAAGGAAGCCCAAGATGACAGATTITGATCGTTTTAAAGTTATGAAGGCCAAAGAAATG 350
QY 372 agaacagaataatcaagaatgaagttaagaagcttcaaaagcagctctcctgaaagct 431
Db 361 AGGAACAGATAATCAAGATGAAGTTAGACAGCTTCANAGGCAGCTCTCTCTGAAAGCT 420
QY 432 tctcccaaaaaagcaccctggtactaagggttactgctgctgctgctgctgctgctgct 491
Db 421 TCTCCCAAAAAAGCACCTGGTACTAAGGTA-----CTGCTGCTGCTGCTGCTGCT 474
QY 492 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 551
Db 475 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 534
QY 552 ccagcccaagagttcctgcccagaagccacagagccagaagcagcgcctgctcctcaaaa 611
Db 535 CCAGCCCAAGAGGTTCCTGCCAAGAACCCACAGGCAGAAAGCAGCGCTGCTCTCCAAA 594
QY 612 gctcagaaggtcaaaaagctccagcccaagaaacacctgctcctcaagcagctgcaag 671
Db 595 GCTCAGAAGGTCAAAAAGCTCCAGCCCAAGAAAGCAGTGTCTCCAAAGCATCTGCAAG 654
QY 672 aaagca-taagtgcacatcataaaagtaataaagttcttttgacctgtt 722
Db 655 AAAGCATTAAAGTGGCAATCATAGAAGTAAAGAGTTCTTGGTGAAGCTGTT 705

